DEPARTMENT OF THE ARMY TECHNICAL MANUAL

TM 11-673

GENERATION AND TRANSMISSION OF MICROWAVE ENERGY

DEPARTMENT OF THE ARMY • *JUNE 1953*

United States Government Printing Office
Washington : 1953

DEPARTMENT OF THE ARMY
WASHINGTON, D. C., *18 June 1953*

TM 11–673 is published for the information and guidance of all concerned.

[AG 413.44 (6 Mar 53)]

BY ORDER OF THE SECRETARY OF THE ARMY:

J. LAWTON COLLINS
Chief of Staff, United States Army

OFFICIAL:
WM. E. BERGIN
Major General, USA
The Adjutant General

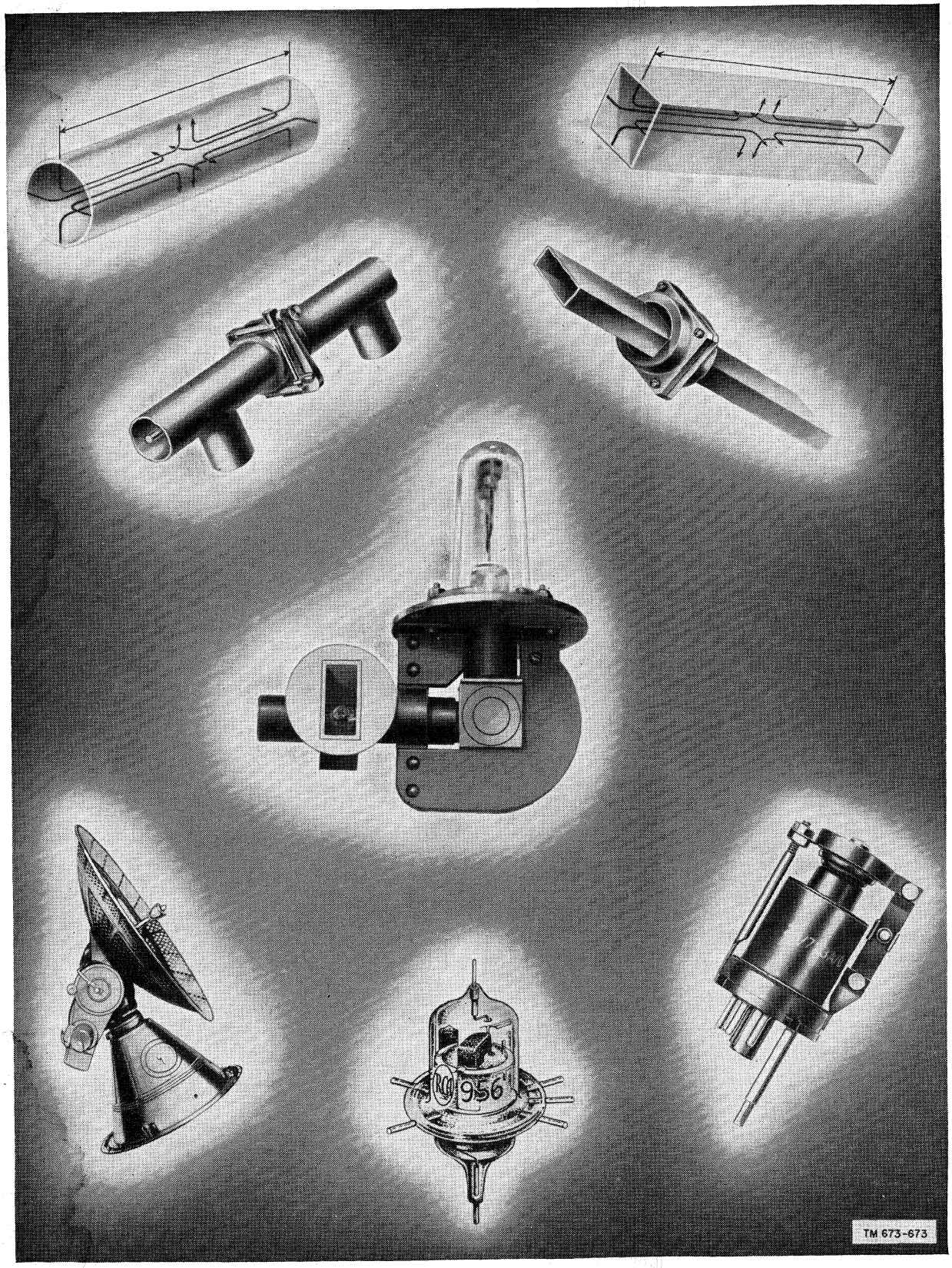
DISTRIBUTION:
Active Army:
Tech Svc (1); Tech Svc Bd (1); AFF Bd (ea Svc Test Sec) (1); AFF (5); AA Comd (2); OS Maj Comd (5); Base Comd (3); MDW (5); Log Comd (2); A (5); CHQ (2); FT (1); USMA (2); PMS & T 11 (1); Dep (2) except 11 (20); RTC (2) except 11 (100); POE (1), OSD (3); Mil Dist (1); T/O & E's: 11–57N (2); 11–95A (2); 11–500A, RR (2).
NG: Same as Active Army.
Army Reserve: Same as Active Army.
For explanation of distribution formula, see SR 310–90–1.

ELECTRONIC FUNDAMENTALS SERIES

The manuals on electronic fundamentals form a progressive series of educational texts which present the theory and application of electronics for the military services. The series starts with the basic electrical fundamentals and extends to the most recent technical concepts, as applied to telegraphy, telephony, radio, loran, facsimile, radio direction finding, radar, meteorological radio, television, and other military equipments.

CONTENTS

	Paragraphs	Page
CHAPTER 1. INTRODUCTION	1–12	1
2. PROPAGATION AT MICROWAVE FREQUENCIES		
Section I. Behavior of microwaves	13–17	9
II. Effects of atmospheric conditions	18–20	12
III. Effect of reflected signal	21–28	16
IV. Line-of-sight limitations	29–39	22
V. Other limitations on maximum range	40–44	31
CHAPTER 3. MICROWAVE TRANSMISSION SYSTEMS		
Section I. Coaxial line	45–48	35
II. Waveguides	49–57	39
III. Tuning devices and related equipment	58–62	53
IV. Couplings, joints, bends, and terminations	63–71	63
V. Other waveguides	72–74	75
VI. Practical considerations	75–78	77
CHAPTER 4. CAVITY RESONATORS		
Section I. Theory	79–85	81
II. Applications of cavity resonators	86–90	91
CHAPTER 5. MICROWAVE OSCILLATORS		
Section I. Oscillators in general	91–95	101
II. Some oscillators using conventional tubes	96–100	107
III. Klystron oscillators	101–105	119
IV. Cavity magnetrons	106–110	126
CHAPTER 6. MICROWAVE RECEIVERS		
Section I. Radar	111–116	135
II. Communications	117–121	151
CHAPTER 7. MICROWAVE ANTENNAS		
Section I. Transmission line terminations	122, 123	158
II. Radiation patterns	124–129	160
III. Use of reflectors	130–135	168
IV. Reflector feed systems	136–138	174
V. Metal lenses	139, 140	178
VI. Scanning and stabilization	141–148	180
APPENDIX I. MATHEMATICAL CONSTRUCTION OF PROFILE DIAGRAM	------	189
II. WAVEGUIDE MATHEMATICS AND EQUIVALENT CIRCUITS OF AMPLIFIER AND QUARTZ CRYSTAL	------	191
III. CHARACTERISTIC WAVELENGTH EQUATION OF RECTANGULAR CAVITY RESONATOR	------	199
INDEX	------	200



Frontispiece.

CHAPTER 1

INTRODUCTION

1. Definition of Microwaves and Scope of Manual

Radio waves at frequencies above 1,000 mc (megacycles) per second are called microwaves. This manual discusses the behavior of microwaves and describes most of the specialized components used in microwave equipment. Anyone who is familiar with vhf (very-high frequencies) will understand the similar behavior of microwaves in space. The circuits and components used in microwave equipment, however, are very different from those of lower-frequency equipment. For example, the most common transmission system at microwave frequencies is neither the coaxial line of vhf frequencies nor the two-wire line of lower frequencies. Instead, a hollow conductor called a waveguide is used.

2. History of Microwaves

a. In 1936, two papers were published that dealt with hollow waveguides—Hyper-Frequency Wave Guides—General Considerations and Experimental Results, by G. C. Southworth, and Transmission of Electromagnetic Waves in Hollow Tubes of Metal, by W. L. Barrow. Both papers discussed the hollow metal tubes that take the place of conventional transmission lines in operation at microwave frequencies. By 1940, a successful cavity magnetron had been developed in Great Britain, and it was possible, for the first time, to generate substantial amounts of power at microwave frequencies. During that year, samples of the cavity magnetron were brought to this country, and research and development work in the microwave field was begun in the Radiation Laboratory at Massachusetts Institute of Technology. Much of the early work at the Radiation Laboratory was directed toward the design of airborne microwave radar equipment, since operation at microwave frequencies permits the use of relatively small antenna structures. The term *radar* actually is a colloquialism for microwave equipment, although it refers to but one type, that which is used to *see* electronically, by means of a transmitted radio wave that reflects from the seen object, the reflection picked up by a receiver and translated into some indication of the range of the object, its bearing, and, if necessary, its elevation angle. In the beginning, however, the first practical radar equipment operated in the vhf range, and the radar used in the early days of World War II operated at those frequencies.

b. The value of radar was not recognized generally during the early years of the war, because its development was one of the successfully guarded secrets of war preparation. Radar equipment was installed on battleships, submarines, and in airplanes often against the wishes of commanding officers. The Commanding Officer of the British battleship, H. M. S. Hood, in the disastrous fight with the German Bismarck, chose the range given by an optical rangefinder in preference to that of his radar, and failed to hit his opponent. The Commanding Officer of the Bismarck used a radar range and hit the Hood on the first salvo. Although the radar used in the Battle of Britain was operated at vhf frequencies, its effects—vision through fog, clouds, and darkness, accurate ranges and tracking—all were present, and Hitler's bombers took a beating they never expected. The introduction of microwave techniques not only sharpened these qualities; it reduced equipment size, lightened its weight, and extended the scope of radar to the point where it became one of the really decisive factors of the war.

c. Until 1942, the effectiveness of German submarines in the North Atlantic was neutralized to a large extent by the use of airborne antisubmarine radar operating at a frequency of about 200 mc per second. By that time, however, many German submarines were equipped with listening receivers operating at the radar frequency. A submarine with a directional antenna could determine the direction of an Allied antisubmarine aircraft, and a rough estimate of range could be made from the

strength of the received signal. The effectiveness of Allied antisubmarine operations dropped to a low at this time, since a submarine, warned of impending attack, invariably dived before the aircraft was in a position to drop depth charges. The effect of the listening receiver was countered by the aircraft for a time by the use of an attenuator inserted between the radar transmitter and the transmitting antenna. During the final phase of an attack, the radar operator reduced the radiated signal level by adjustment of the attenuator. At the submarine, the operator of the listening receiver noted a *decrease* in the signal strength and concluded that the aircraft was going away when it was, in fact, approaching for an attack. This simple trick accounted for the destruction of a number of submarines, but once the secret became known the submarines no longer waited on the surface. When microwave radar was introduced, the Germans thought for a long time that vhf radar had been replaced by some sort of infrared equipment (operating on heat from the submarine), and they made no effort to develop microwave listening receivers. The result was another sharp increase in the effectiveness of antisubmarine operations. By 1943, most of the vhf airborne radar had been replaced by microwave equipment operating at a wavelength of about 10 centimeters.

d. During the next 2 years, 1943 to 1945, most vhf radar was displaced by microwave equipment, and new types of radar were developed. Among these may be mentioned blind bombing equipment and GCA (ground-controlled approach) equipment which permits ground operators to direct an aircraft to a safe landing under conditions of zero visibility. Neither high-altitude blind bombing equipment nor GCA equipment was practical before the advent of microwave radar, because, in both cases, antenna structures having the required directivity are prohibitively large at vhf frequencies.

e. Wartime development work in the microwave field was directed primarily toward improvement of radar equipment of various sorts. Since the war, however, various types of microwave communications equipment have come into use. Most of these are microwave relay installations which handle telegraph, telephone, or television signals. Figure 1 shows a telephone relay installation designed for short-distance work—across rivers, for example. Figure 2 shows a microwave television relay installation. Operation at microwave frequencies offers two major advantages—the great directivity of microwave antennas permits the use of relatively low-power transmitting equipment, and the great bandwidth available at microwave frequencies affords tremendous traffic-handling capacity. In both respects, mocrowave

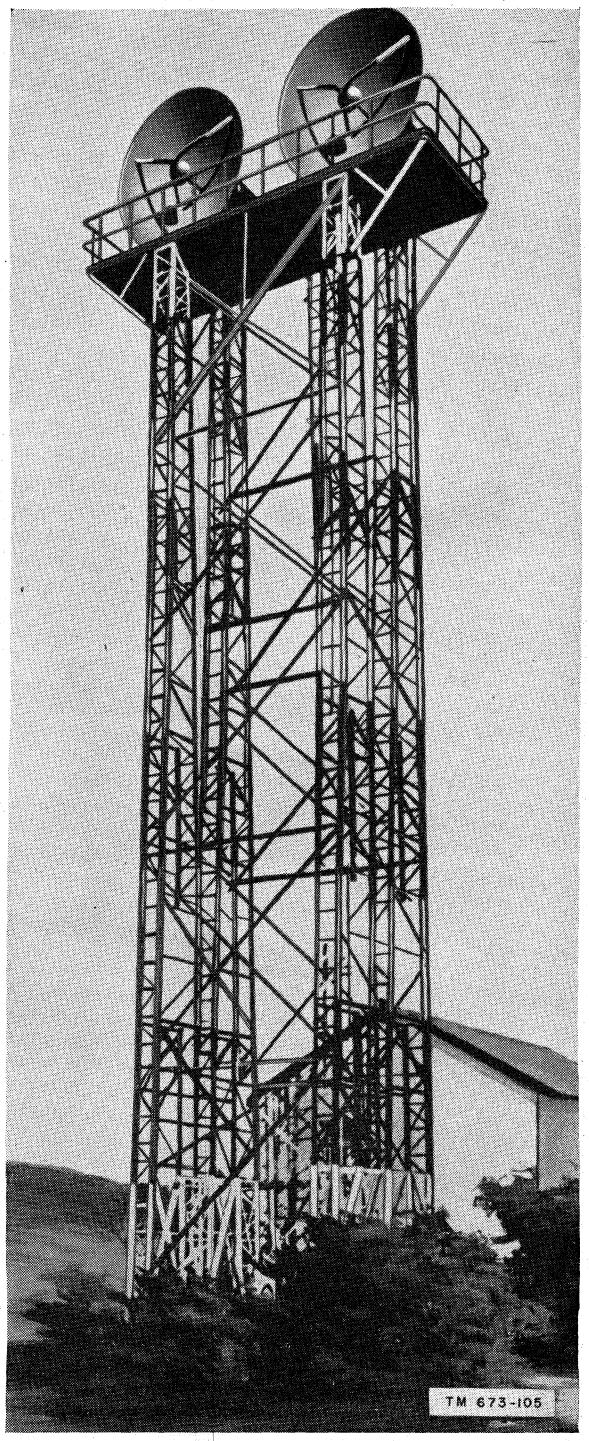


Figure 1. Short-range microwave telephone relay installation.

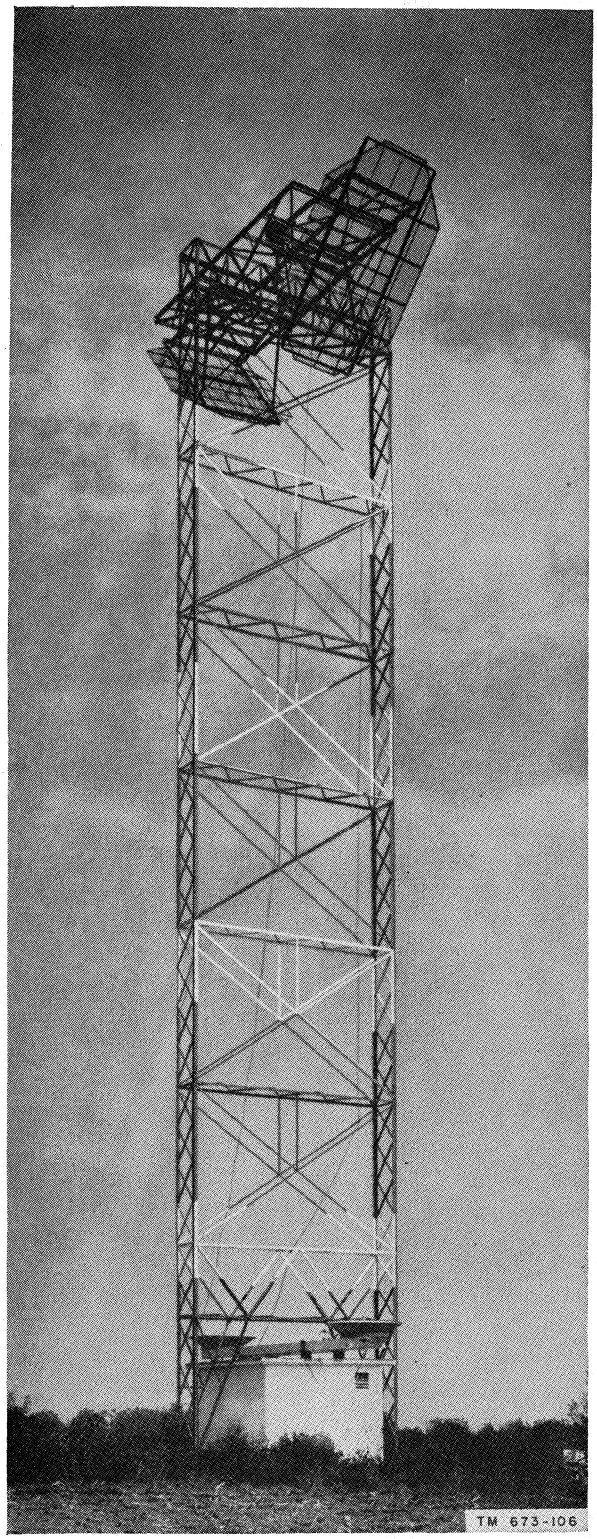


Figure 2. Microwave television relay installation.

equipment is superior to vhf equipment. A single low-power telephone relay installation operating at microwave frequencies, for example, is able to handle more than a thousand simultaneous conversations.

3. Special Aspects of Microwaves and Microwave Equipment

The difference between microwave theory and ordinary circuit theory may be compared to the difference between d-c (direct-current) and a-c (alternating-current) circuit theory. After studying d-c theory, it is necessary to acquire a new point of view in order to understand a-c theory, with concepts like phase, frequency, and reactance which bear little apparent relation to d-c theory. Conventional radio circuit theory is based on voltages and currents, but much of microwave theory is based on electromagnetic fields. This actually is not a great change in thinking. To begin with, the concept of electromagnetic fields forms the basis of all antenna theory. Also, once the operation of waveguides and their associated components is grasped, they may be considered from the familiar viewpoint of transmission line theory. The standard concepts of impedance matching and standing wave ratio apply to waveguide installations just as they apply to conventional transmission line systems. Because of this, there is a close similarity between microwave systems and vhf systems, and this becomes apparent as soon as the fundamentals of waveguide theory have been digested. Just as a-c theory is a logical extension of d-c theory, so microwave practices and techniques follow easily and simply from those of conventional radio.

4. Standard Waveguides

Figure 3 shows two standard types of waveguides, rectangular and circular. At this point, it is necessary to discard the concept of the operation of a transmission system in terms of voltages and currents. Instead, anyone understanding electric and magnetic fields will be close to an understanding of the microwaves that travel *inside* the waveguide. There is simply a change in the terms and the principles used in thinking about transmission systems. Once this change in approach is accepted, a description of waveguides and other components of microwave transmission systems will be grasped easily.

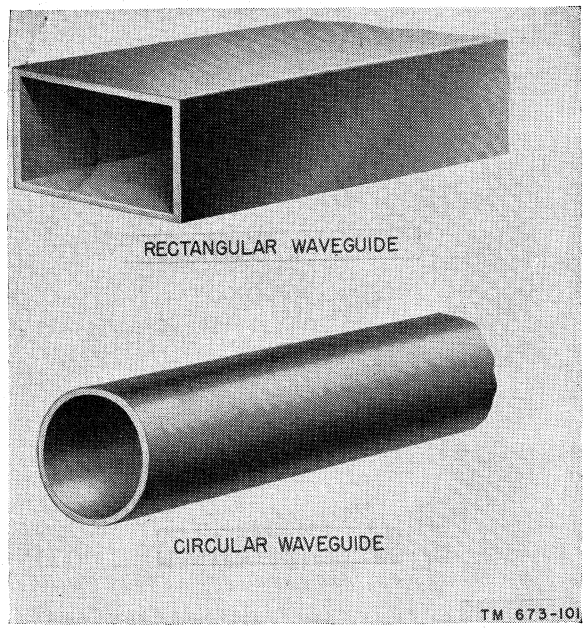


Figure 3. Hollow waveguides.

5. Dimensions of Microwave Components in Relation to Wavelength

a. The dimensions of most microwave components are nearly proportional to the wavelength at which they operate, and therefore the size of the equipment varies with the wavelength of the signals. For obvious reasons, microwave systems usually are designed for operation in one of the three standard bands in the following tabulation, and components normally are available in sizes adapted to the three wavelengths.

Bands	Frequencies (mc per second)	Wavelengths	
		(cm)	(in.)
S-band	Near 3,000	10	4
X-band	Near 10,000	3	1. 2
K-band	Near 27,000	1. 1	. 44

b. The three standard sizes of a bent section of waveguide are shown in figure 4. The largest section is used for the S-band, the middle size for the X-band, and the smallest size for the K-band. The flange fittings on the end of each section are used for connection to adjacent pieces of waveguide.

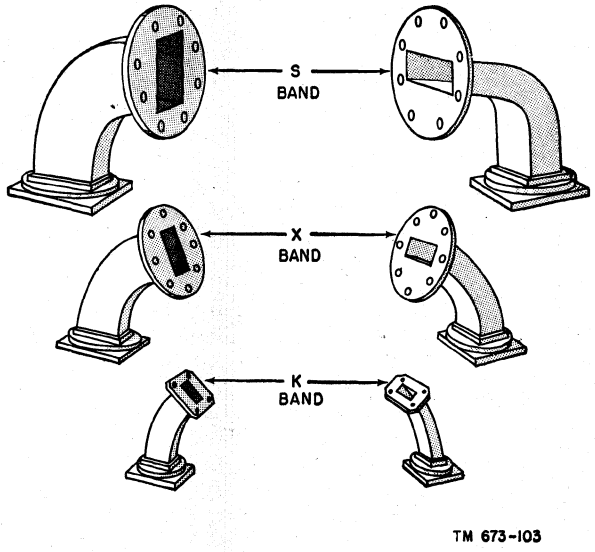


Figure 4. Standard sizes of one type of waveguide bend.

6. S-, X-, and K-Band Usages

Although signals in the S-, X-, and K-bands behave in similar ways, the effects of rain, snow, and fog are more pronounced in the higher-frequency X- and K-bands. For this reason, early-warning radars which must pick up targets at long ranges under all weather conditions usually are operated at S-band frequencies. Most airborne radars operate at X-band frequencies to take advantage of the weight saving afforded by the reduction in size of microwave components, since the size of these is usually proportional to the wavelength. Antiaircraft tracking radars

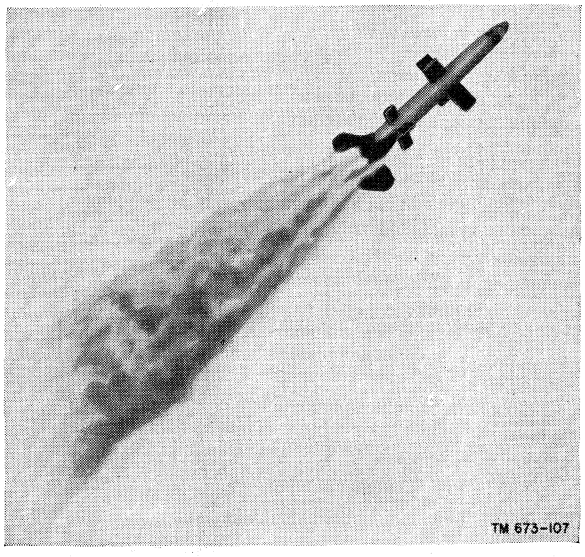


Figure 5. Guided missile radar.

also operate at X-band frequencies, sacrificing long-range, bad-weather performance to obtain the greater tracking accuracy resulting from the sharper beam obtainable in the X-band. The K-band is reserved for special applications, such as radars, where very light weight and small antenna size are more important than reliable long-range operation. A typical example of K-band equipment is the midget homing radar used in a guided missile. Figure 5 shows a guided missile containing a radar of this sort.

7. Bandwidth of Microwave Systems

One of the advantages of operation at microwave frequencies is the great bandwidth of microwave systems or, in other terms, the tremendous amount of information that can be carried by a single microwave system. When an r-f (radio-frequency) carrier of frequency f_c is amplitude-modulated by an audio signal of frequency f_a, the resulting modulated carrier consists of three distinct r-f signals. The first of these is the original carrier. The others, at frequencies f_a above and below the carrier, are called the *upper and lower side bands*, respectively. If the carrier is frequency-modulated, there are additional side bands at frequencies $(f_c+2f_a),(f_c-2f_a),(f_c+3f_a),(f_c-3f_a)$, and so on. The r-f section of any transmitter is, necessarily, tuned to the carrier frequency f_c and will pass only signals having frequencies within about 5 percent of f_c. (The limit is, of course, different with different types of equipment.) In practice, f_a often is limited to values such as $.01f_c$. It follows, then, that higher modulating frequencies can be used when the frequency of the carrier is increased. Consider, for example, the 96 channels at frequencies of 550 kc (kilocycles), 560 kc, 570 kc, and so on up to 1,500 kc. Operating on these channels, 96 transmitters can carry 96 separate voice programs. It is possible, however, to operate 96 low-power r-f amplifiers at these frequencies and modulate each with its own voice program. The outputs of the 96 modulated amplifiers then can be combined (by adding them together) into a single rather complex signal containing frequencies that range from slightly below 550 kc to slightly above 1,500 kc. This complex signal then can be used to modulate a microwave transmitter. It is, therefore, possible to carry all of the 96 voice programs in the *entire* broadcast band on a single microwave carrier. At the receiving end of the microwave system, routine detection of the incoming signal provides the complex signal (containing the 96 modulated carriers) at a high level with a negligible noise background. Then, 96 simple receiver-type circuits select the separate modulated carriers, and the outputs of these circuits are detected in the usual way to obtain the 96 voice programs—each on a separate line. Thus, through the use of more refined techniques (basically the same as the technique described above) a single microwave relay installation is enabled to handle over a thousand telephone conversations at the same time.

8. Directivity of Microwave Systems

Another advantage of operation at microwave frequencies is the great directivity that can be obtained with relatively small antenna systems. In general, the sharpness of an antenna pattern depends on the size of the antenna. If two transmitting antennas are operated at the same frequency, the larger one ordinarily will radiate a narrower beam than the smaller one. Similarly, if two receiving antennas are operated at the same frequency, the larger one will have greater directivity than the smaller one. It turns out, however, that two antennas operating at different frequencies will have similar radiation or similar reception patterns if their dimensions in wavelengths are similar. As an example, an antenna array 10 feet high and 20 feet wide operating at a wavelength of 1 meter (at a frequency of 300 mc per second) has essentially the same radiation pattern as an array 20 feet high and 40 feet wide operating at a wavelength of 2 meters (at a frequency of 150 mc per second). Thus, microwave antennas of small size may achieve the same radiation pattern as the larger ones used in vhf systems.

9. Relative Antenna Sizes

One type of early-warning radar, which was used extensively before microwave equipment was developed, operated at a frequency near 700 mc per second and used an antenna array about 6 feet wide by nearly 6 feet high. This array (fig. 6) is a heavy and awkward amount of gear to swing around on the top of an antenna tower. The machinery necessary to turn it and the structure required to support it are necessarily large and expensive. With microwave equipment, a similar pattern can be obtained at S-band fre-

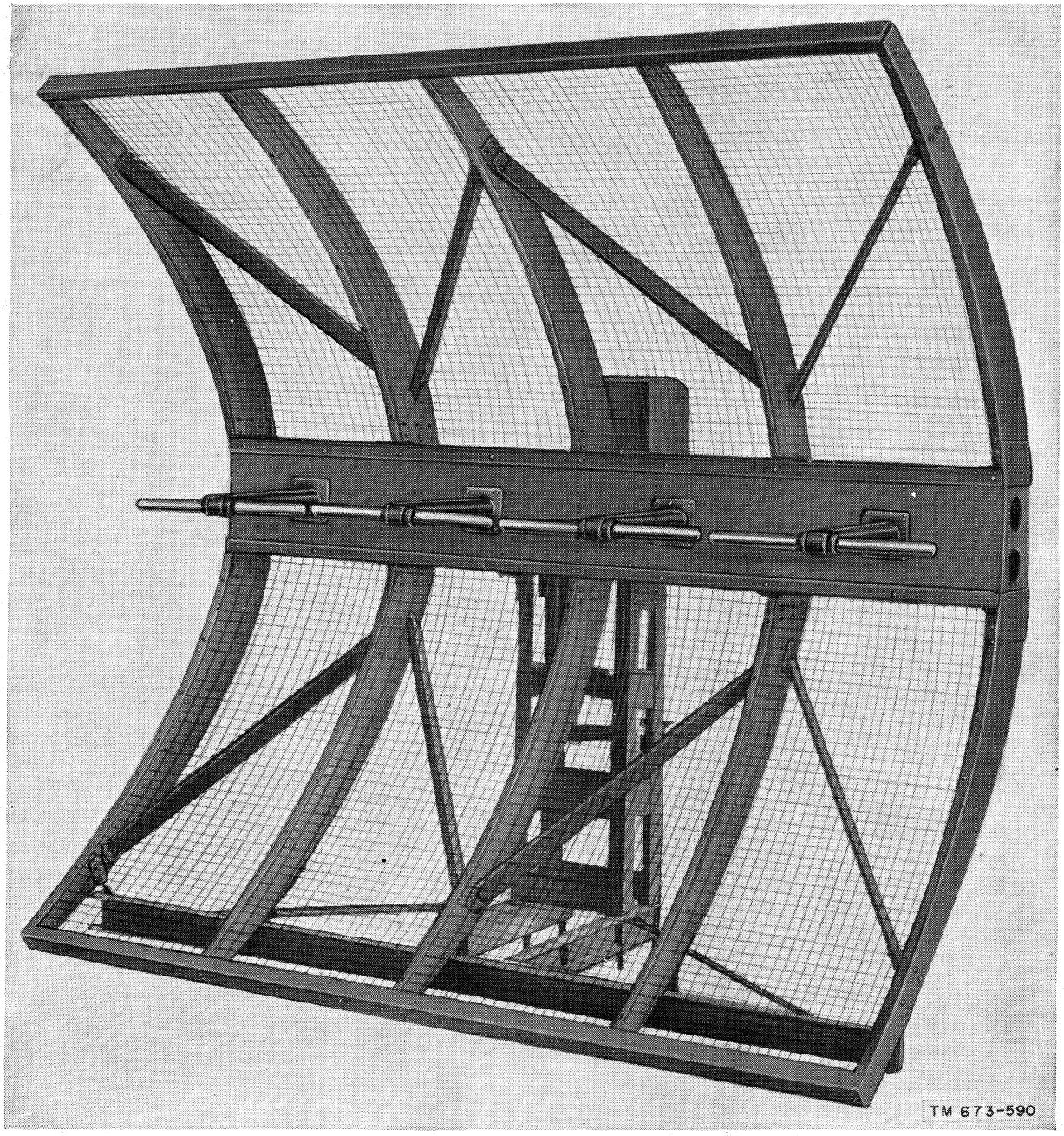


Figure 6. Vhf early-warning radar antenna.

quencies (around 3,000 mc) with an antenna approximately 16 inches wide by 12 inches high, or with a circular structure slightly over 6 inches in diameter. In practice, an S-band antenna is rarely as small as this, but it is obvious that the saving in weight and size may be considerable.

10. Usual Microwave Antennas

Antenna sizes for airborne radars are generally about 30 inches in diameter for the S-band, and anywhere from 8 to 30 inches for the X-band. A present-day early-warning radar operating in the S-band uses an antenna about 3 feet by 10 feet, approximately one-quarter the size of the vhf radar mentioned earlier, but its efficiency is so increased that there is no comparison between its performance and that of its vhf forerunner. A representative ground radar used to control anti-aircraft guns operates in the S-band with an antenna diameter of about 60 inches (fig. 7.)

Generally speaking, microwave radars in use today operate with smaller antennas and sharper beams than the vhf radars in use in the early years of World War II. The advantages of smaller antennas are obvious both in airborne equipment and in shipboard and mobile installations, where space is limited. The sharper beam of a microwave antenna allows more accurate determination of target direction, since echo strength changes more rapidly when the beam is moved off the target. At the same time, the sharper beam concentrates the radiated energy, thus delivering more power to the target in the case of radar, or to the receiving antenna in the case of a point-to-point communications system.

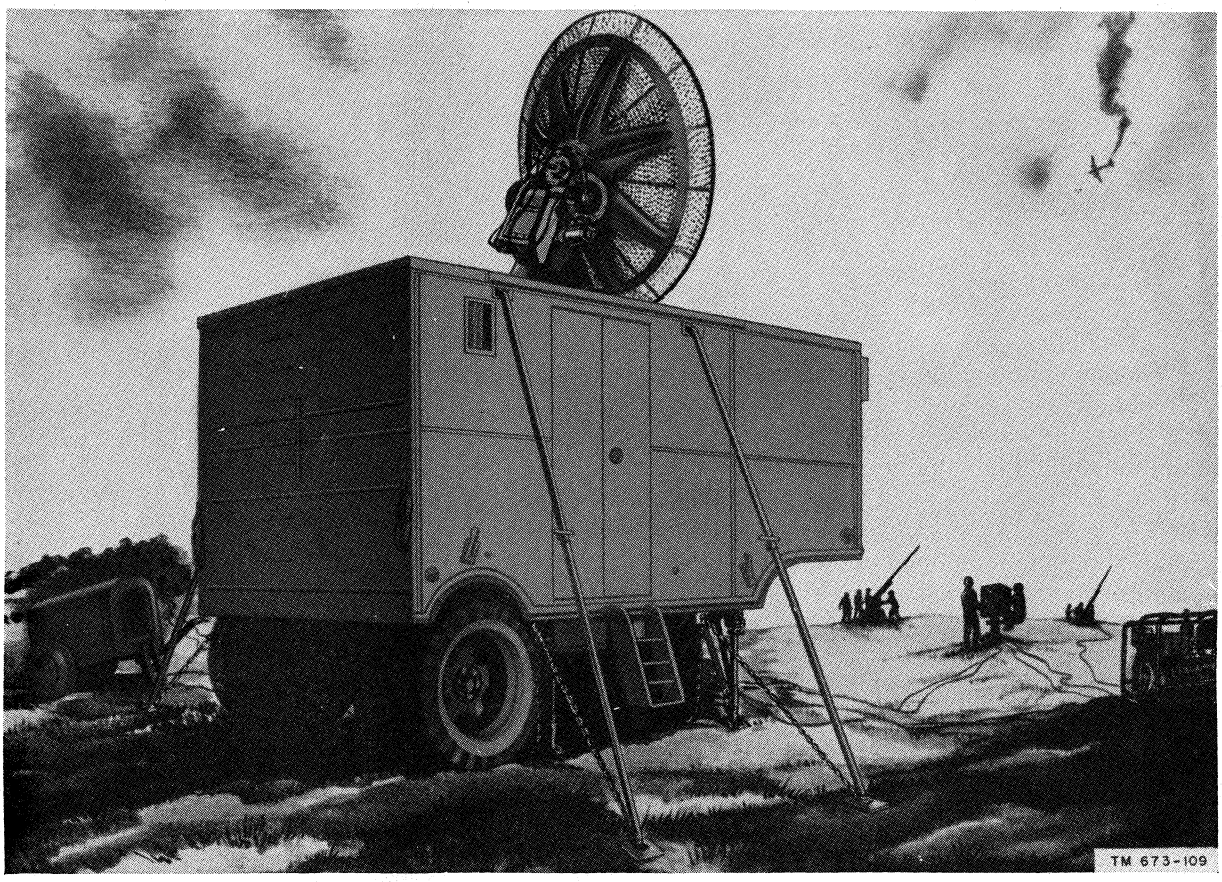


Figure 7. Typical radar for antiaircraft gunfire control.

11. Summary

a. Radio waves at frequencies of more than 1,000 megacycles per second are called microwaves. The sizes of most microwave components depend on the wavelength at which they operate, and most microwave equipment is designed for operation in one of three standard bands for which components are widely available. These are the S-band (wavelength about 10 cm, frequency about 3,000 mc per second), the X-band (wavelength about 3 cm, frequency about 10,000 mc per second), and the K-band (wavelength about 1.1 cm, frequency about 27,000 mc per second).

b. Practical use of microwaves dates back to the middle of World War II. Wartime application of microwaves was limited to radar equipment, but postwar developments include microwave communications equipment—particularly television relays and multiple-channel telephone and teletype relays.

c. The important merits of microwave operation are the great bandwidth available (which allows a single microwave relay installation to handle as many as a thousand simultaneous telephone conversations) and the great directivity obtainable from relatively small antenna structures.

12. Review Questions

a. Of what is the most common transmission system used at microwave frequencies composed?

b. Conventional radio circuit theory is based on voltages and currents. On what is microwave theory based?

c. Name the three standard bands used in microwave systems, and give the frequencies near which they operate.

d. Name a standard equipment use for each of the three microwave bands.

e. Is the bandwidth of a microwave system great or small?

f. Why is great bandwidth desirable?

g. In what way is the size of microwave equipment related to frequency?

h. Can microwave antennas of small size achieve the same results as the larger ones of vhf systems?

CHAPTER 2
PROPAGATION AT MICROWAVE FREQUENCIES

Section I. BEHAVIOR OF MICROWAVES

13. Importance of Propagation Characteristics

a. It usually is possible to obtain reliable operation over distances up to 100 miles with radio systems operating at frequencies below 30 mc per second, but this is not true with systems operating at microwave frequencies. Even with a powerful transmitter and a good antenna system, one cannot be sure of delivering a usable signal at a specified point. In some cases, for example, a change of 50 feet in the height of the transmitting antenna may provide excellent operation where *no operation* was possible before. In other cases, communication between two points may be possible only under abnormal weather conditions. Such behavior results from variation in the *propagation characteristics* of the path joining the transmitter to the desired receiving point.

b. Given two points, the microwave propagation characteristics of the path between them—that is, the way in which microwave energy travels from one point to the other—depend on such factors as the height of each end of the path, the length of the path, the nature of the terrain between the two points, the weather conditions along the path, the particular frequency used, and so on. Because of this variation, the propagation characteristics of the path between the transmitting and receiving antennas of a point-to-point communications system, or of the paths between the antenna and possible targets, in a radar installation, are of the greatest importance.

c. It is necessary to be able to predict the propagation characteristics of any given path, so that suitable paths can be chosen and equipment can be installed at sites that will permit the most reliable operation possible.

14. Four Paths of Radio Waves

a. Four different paths may be involved in the transmission of energy, at radio frequencies, from one point to another. These are the *direct wave*, the *reflected wave*, the *surface wave*, and the *sky wave*. In any given instance, one, two, three, or all four types of waves may be present, all issuing from the same radiating source. This concept is suggested in figure 8, where only the direct and reflected waves are represented.

b. Most microwave installations make use of the direct wave (signal) only, although in some cases the reflected wave is also important. This characteristic of microwave systems may be contrasted with the behavior of similar installations operating in the vhf range, where the direct and reflected waves are usually (but not always) of equal importance.

c. The other two waves, the surface wave and the sky wave, are mentioned only because of their importance at low and medium frequencies. The surface wave is seldom important in either microwave or vhf operations, and the sky wave is absent at frequencies higher than 60 mc per second.

15. Microwaves

a. Direct Wave. The direct wave from one point to another takes its name from the fact that it travels directly—that is, with no reflections—between the two points. Figure 8 shows the paths followed by the direct and the reflected waves traveling from the antenna of an early-warning radar to the aircraft that is under observation. As already noted, the success or failure of a microwave system usually depends on the presence or absence of the direct wave, since the other signals ordinarily are absent or unimportant. The first step in predicting the propagation characteristics of a given path is, therefore, consideration of the direct wave.

b. Reflected Wave.

(1) The term *reflected wave* is applied to a wave that has been reflected from the

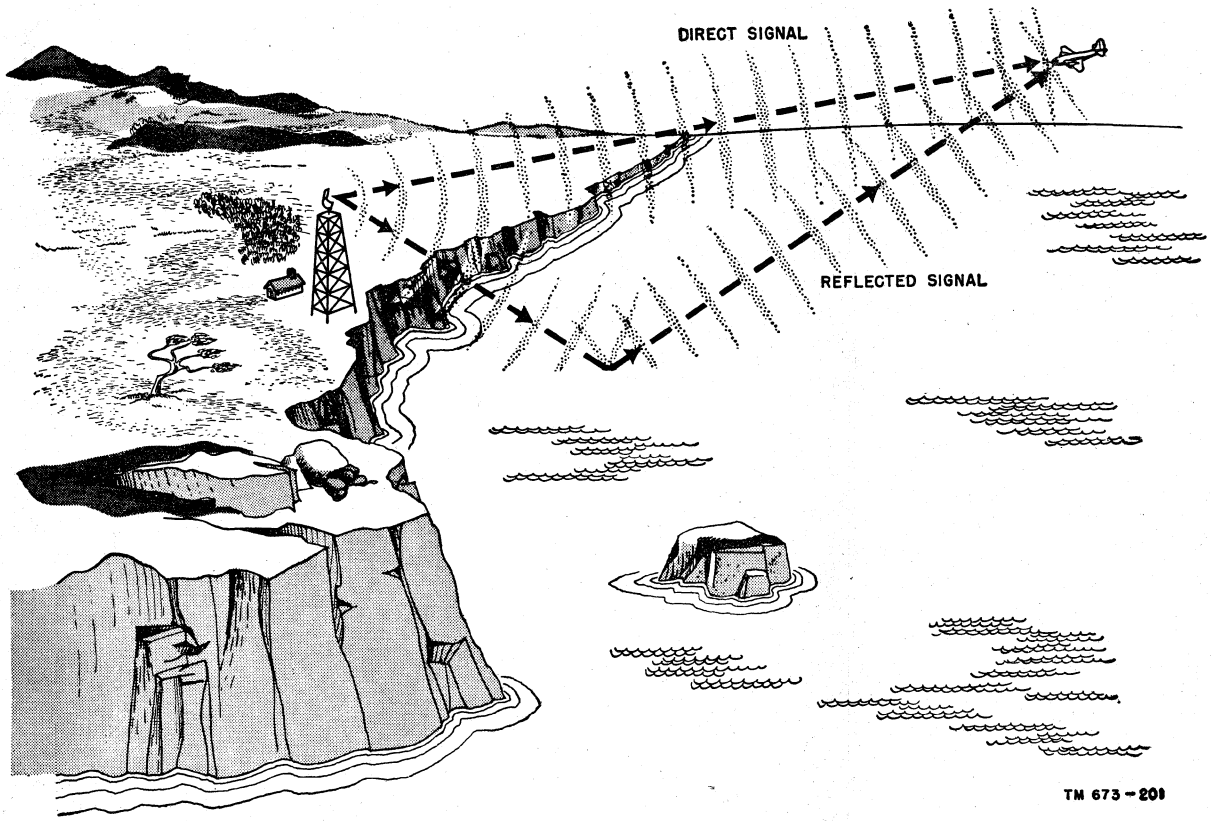


Figure 8. Direct and reflected waves.

land or water surface of the region between the transmitting and receiving points. Figure 8 shows the path of such a wave. There is usually a path of this sort, but there can be no reflected signal unless the transmitting antenna radiates energy along the path. In point-to-point ground communications, the beamwidth of the antenna is such that ground reflections are always present.

(2) A somewhat similar limitation holds for propagation in the reverse direction—from the aircraft to the antenna. Since the aircraft radiates an echo in *all* directions, a reflected signal always will be delivered to the antenna. This signal, however, has no effect on the performance of the system unless the antenna is able to pick up energy from the direction in which the wave arrives.

(3) Another factor that affects the reflected wave is the nature of the reflecting surface. If the surface is smooth, any energy arriving along the reflected-wave path will be reflected in an orderly way, and at least a large part of it then will travel to the target (A of fig. 9). If, however, the surface is rough, the energy that reaches it will be reflected in a multitude of different directions and only a small portion of the energy will continue along the reflected-wave path to the target, as in B. If the reflecting surface is rough, the reflected signal at the target will be so much weaker than the direct signal that it may be neglected. Generally speaking, water acts like a smooth surface at microwave frequencies and land acts like a rough surface.

(4) Although this discussion has been in terms of a target—the aircraft shown in figure 8—the same considerations hold when the path ends at a receiving antenna.

(5) The importance of the reflected wave lies in the fact that the effective signal at the target—or at the receiving antenna—is

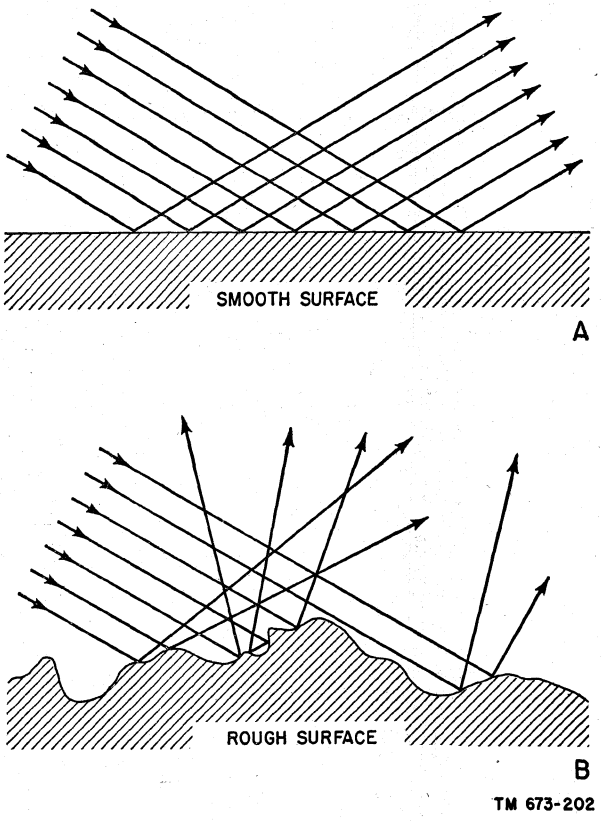


Figure 9. Reflection from smooth and rough surfaces.

the sum of the direct and reflected waves. Like any other alternating waves, a-c or r-f, these signals must be added with due regard to the phase difference (if any) between them. A phase difference between the two waves often is introduced by the difference in the lengths of their paths. This difference may be anything from a small fraction of a wavelength to many wavelengths. When the path difference is an odd number of wavelengths, the two signals arrive at the target or receiving antenna in phase. This is true when the wave is reflected at small angles of incidence, causing a 180° phase reversal. For horizontal polarization, the phase reversal is nearly 180°, irrespective of the frequency and magnitude of the grazing angle. This is true also of vertical polarization for almost all point-to-point ground communications. With vertical polarization at grazing angles in access of, say, .5°, the phase reversal is less than 180°, depending on the frequency and the grazing angle.

16. Signals Present at Lower Frequencies

a. Sky Wave. A sky wave is one that has been reflected from the ionosphere, a region that extends from an altitude of 40 or 50 miles to the upper limit of the atmosphere. The mechanism of ionosphere reflection of radio signals differs somewhat from that of ordinary reflection by a surface, described in paragraph 15*b*. Also, ionosphere reflection diminishes as the signal frequency increases, and, therefore, it has no bearing on microwave propagation.

b. Surface Wave. The surface wave travels along the surface of the earth or water. It is not present unless the transmitting antenna is within a few wavelengths of the ground and, since it remains close to the surface, it cannot be picked up unless the receiving antenna is also low. The surface wave signal is important at low frequencies; most daytime broadcast reception is by surface wave. It is relatively unimportant at vhf and may be neglected completely at microwave frequencies.

17. Curvature of Microwave Paths

a. The paths followed by microwave and vhf signals, whether direct or reflected, usually are slightly curved. The signals travel through the atmosphere at speeds that depend on temperature, atmospheric pressure, and the amount of water vapor present in the atmosphere. Generally, the higher the temperature, the faster the signal; the lower the atmospheric pressure, the faster the signal; and the less the amount of water vapor present, the faster the signal. The net result of these influences is that the signal speed changes with altitude, and, under normal atmospheric conditions, the variation is a small and uniform increase in the speed of the signal with increasing altitude. This causes all paths to curve slightly downward, as shown in figure 10.

b. To understand the reason for this curving downward, imagine a surface called *a wave front* such as the one represented by line AB (fig. 10), where A is higher in altitude than B. A wave

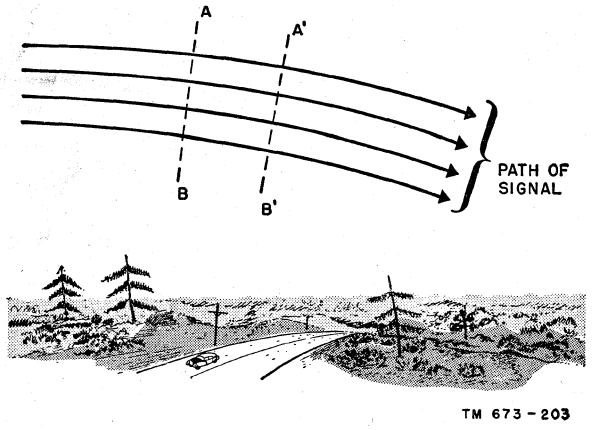


Figure 10. Microwave path curvature.

front is a surface, not necessarily flat, chosen in such a way that the *phase of the signal is the same at all points on the surface.* The signal always travels in a direction perpendicular to the wave front. Now, as the wave front, AB, moves with the signal, it reaches a position, A′B′, after a short time interval. The speed at A and A′ is faster than the speed at B and B′, since A and A′ are at a greater altitude. Therefore, in a given time, the upper part of the wave front moves farther than the lower part, and the wave front tips slightly forward as it moves along. Since the path followed by the signal is always perpendicular to the wave front, it curves slightly downward as the wave front tips.

Section II. EFFECTS OF ATMOSPHERIC CONDITIONS

18. Standard Propagation

Reference has been made in paragraphs 13–17 to *normal atmospheric conditions.* These are described in this paragraph, and the more common abnormal conditions are described in the paragraph following.

a. Index of Refraction.

(1) Both light waves and radio waves travel through a vacuum with a speed nearly equal to 330 yards per microsecond (a microsecond is 1/1,000,000 of a second). In more common units, they travel nearly 186,000 miles per second (these are statute miles of 5,280 feet). In almost any other medium, the velocity of the waves differs from their velocity in a vacuum, and signals of different frequencies travel with different velocities. Visible light, for example, travels through water about three-fourths as fast as through a vacuum, and radio waves travel through water only one-ninth as fast as through a vacuum. (Radio waves are, of course, greatly attenuated, or diminished, by water and for practical purposes it may be considered that they do not travel through water at all.)

(2) Instead of saying that visible light travels through water with a velocity of about 247.5 yards per microsecond, it is more convenient to say that water has an *index of refraction* of about four-thirds for visible light. The index of refraction of water is simply a number equal to the velocity of light in a vacuum divided by its velocity in water. Similarly, water may be said to have index of refraction of four-thirds for visible light and nine for radio waves, and a vacuum has an index of one for both types of signals. The velocity of a light wave or a radio wave through any medium is then very nearly $330/n$ yards per microsecond, where n is the index of refraction of the medium for signals of the particular wavelength under consideration.

b. Normal Variation of Index of Refraction With Height.

(1) For microwave and vhf signals, the index of refraction of air becomes smaller with decreasing pressure, smaller with decreasing moisture content, and larger with decreasing temperature. All three of these variables—pressure, moisture content, and temperature—decrease with increasing altitude in a normal atmosphere. The changes in pressure and moisture content make the index smaller at greater heights. The change in temperature tends to make the index larger, but it is not sufficiently great to offset the decrease caused by reduction in pressure and moisture content. The net result, under average conditions, is that the index of refraction for microwave signals decreases uniformly with height up to altitudes of approximately 25,000 feet.

(2) Because the velocity of microwave signals increases with height, the signals follow paths that are curved slightly downward, as explained in this paragraph and in paragraph 17.

(3) Although the index of refraction for light waves also varies with height, it changes much less than the index for microwaves, and the paths followed by light waves are nearly straight lines.

(4) Under normal atmospheric conditions, the curvature in the paths of microwave signals is such that the paths appear as straight lines when they are plotted on a diagram in which the curvature of the earth's surface is distorted by using a fictitious value, equal to four-thirds of the actual value, for the earth's radius. (Diagrams thus distorted actually are made in practice, and are explained in pars. 29–39.) Such atmospheric conditions are called *standard conditions*, and propagation (the process by which the wave travels, or the way in which it travels) under such conditions is called *standard propagation*.

19. Anomalous Propagation

Temperature and moisture content are interrelated in such a way that normal changes in these quantities do not alter appreciably the standard variation of the index with height. Because of this, standard conditions generally prevail. However, it is possible to have conditions under which the variation of the index with height is not standard. Such conditions are called *nonstandard conditions*, and propagation under such conditions is called *anomalous propagation*.

a. Light waves are refracted, or bent, when they pass from one medium to another medium having a different density. This physical fact accounts for the phenomenon known by travelers on the desert as a mirage. Thus, an oasis that is actually many miles away appears to be clearly above the horizon. What has happened is that light rays from the distant oasis have been bent around the earth's curvature. This effect may result when the atmosphere is not uniform and is subject to varying composition and density, causing the light rays to travel in a curved path instead of in a straight line. Similarly, radio waves can behave in a like fashion. Both the optical mirage and the radio wave mirage are caused by the bending of waves away from a medium of low density into a medium of higher density. The degree of bending is expressed by the changes in *refractive index*. The atmosphere may provide a curved path by which radio waves may be bent around the earth's curvature; such a curved path for radio waves is called a *duct*.

b. Atmospheric conditions of abnormal variation affect the operation of radio equipment on frequencies greater than 30 mc. In general, the higher the frequency the greater will be the possibilities of a duct causing erratic radio-wave propagation effects. The radar operator using frequencies in the microwave region must be on guard against false readings attributable to atmospheric peculiarities. For this reason, an echo box is best for equipment performance testing instead of a target such as a cliff or a distant mountain; such targets may not be dependable reference points, since abnormal propagation effects because of trapping can give false readings in signal strength as great as 50 db (decibels).

c. The question arises—What are the weather conditions that cause the creation of a duct and the resulting propagation peculiarities? It is well known that the earth's atmosphere is subject to many variations. In general, both the temperature and the moisture content decrease gradually and practically uniformly with height above the surface of the earth. This is the standard condition, and when this condition occurs no ducts are formed and radio transmission is normal. However, normal atmospheric conditions may not be present at all times. In some cases, the temperature first may increase with height and then begin to decrease. This is referred to as *temperature inversion*. In addition, the moisture content of the atmosphere may decrease very sharply close to the earth's surface. When these conditions exist—namely, temperature inversion accompanied by a region of great moisture variation—a duct, which can cause downward bending of radio waves, is likely to be formed.

d. A modified index of refraction curve, usually called an M-curve (fig. 11), is used to predict the probable degree of bending. Changes in M indicate the bending of the radio wave in the atmosphere. A, of figure 11, shows a straight line sloping up gradually to the right without any break. A curve of this nature predicts standard propagation. Conversely, a curve that slopes up and to the left indicates that duct-forming condi-

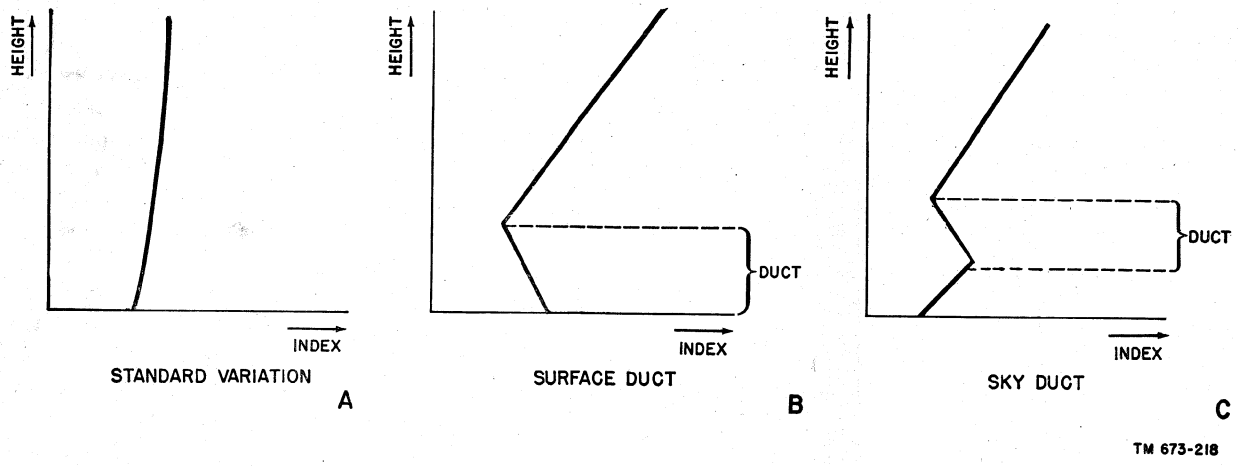


Figure 11. Modified index of refraction curves.

tions prevail. B shows an M-curve that rises in a direction toward the left, which signifies the possibility of a *surface* duct. In C, the curve rises to the right and then breaks to the left. This discontinuity discloses the existence of atmospheric conditions likely to cause an *elevated* duct.

e. One of the weather conditions that causes anomalous effects results from warm, moist air flowing over cold water; the air layers closest to the sea are cooled more than higher levels, and temperature inversion is the result. Since the upper levels are warmer, they can hold more moisture than the lower colder levels and, therefore, the vapor content increases with height. This weather situation may cause a duct to be formed and lead to nonstandard propagation.

f. Another weather situation that may cause the formation of a duct results from surface (radiation) inversions. This condition occurs when the heat is being radiated from the earth's surface faster than it can be dispersed by the lowest level of the atmosphere; the result is that the temperature of the earth's surface is much lower than that of the air, and the temperature inversion which occurs may form a low-level duct. Radiation ducts may be expected over large, flat land areas in the middle and higher latitudes. Desert country is such a region where this type of radiation duct may be encountered. Ducts of this character generally are found at night or in the early morning. In addition to the examples just cited, other weather situations can cause unusual changes in temperature and moisture distribution in the atmosphere; two of these are the flow of cool air over a warmer sea and a descending or subsiding body of air which acts to bring about a warmer, dryer level.

g. As already mentioned, the paths of microwaves in a duct curve downward more sharply than usual. Therefore, they curve downward even when plotted over a modified earth of the type mentioned in paragraph 18*b*. The presence of a duct always increases the operating range of a radar (assuming that the range is limited by the curvature of the earth rather than by available power) and, when the curvature of a horizontal microwave path is the same as the curvature of the earth, the range may be extended by as much as several hundred miles, as shown in B of figure 12. A shows the normal range of the same radar. Both drawings are plotted on a fictitious earth curvature (par. 18*b*) so that the signal paths in standard atmosphere, or *above* the duct, appear as straight lines. Although ducting increases the operating range of a radar, it has a bad effect in that it may decrease the *coverage* of the radar. B shows that the high-altitude coverage of the radar at long range has been reduced. In general, increased range is obtained when both target and radar lie in the duct, whereas the range is reduced when one is in the duct and the other is not. Because of this, an aircraft making an antishipping search by radar should always fly in a duct (when one is present) rather than above it. Since some energy is lost by the signal as it travels over the water, and some energy leaks out the top of the duct, extreme increases in range seldom are encountered except with ducts that are several hundred wavelengths high.

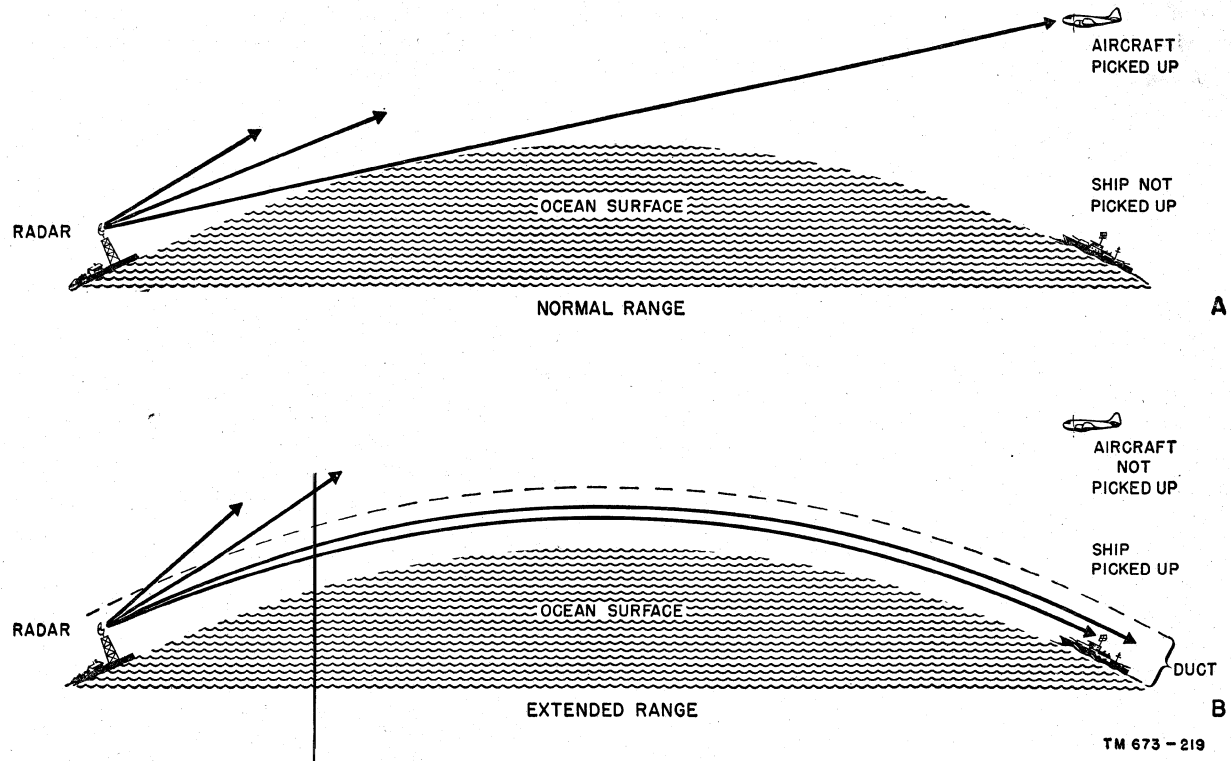


Figure 12. Extension of range by duct.

20. Absorption and Scattering

The propagation characteristics of microwaves already discussed—curved paths and anomalous propagation—also hold for the vhf range of frequencies. Two additional atmospheric effects peculiar to microwaves are absorption and scattering.

a. Absorption. At certain frequencies, the atmosphere will not transmit radio waves but will, instead, absorb most of the energy. The effect has nothing to do with reflection. Instead, the energy is absorbed in precisely the same way that red light is absorbed by a pair of green sun glasses (which transmits only the green light). Certain of the gas molecules, particularly those of oxygen and water vapor, are able to oscillate at these frequencies, and they behave exactly like resonant circuits that absorb energy only at the frequencies to which they are tuned. The effect is not present at frequencies lower than the X-band, and does not become serious until the K-band is reached. Because a number of strong absorption bands are present in the region between .1 and 1 cm, it is likely that there will be no conventional microwave equipment at wavelengths shorter than the present K-band. The next step in reducing wavelength probably will be the development of infrared equipment operating at wavelengths of less than 1 millimeter.

b. Scattering. The effectiveness of a microwave system may be reduced sharply by scattering of the signal. This phenomenon depends on the presence of small drops of water or fine particles of dust, snow, or ice in the air. These act as reflectors, and divert some of the signal from its original path. The direction in which the signal is reflected has little relation to the original direction—except that it is not the same. The greater part of the energy is radiated in directions at right angles to the original path, and some even is reflected back in the opposite direction. As might be expected, the effect is worse at higher frequencies because of the greater size of the particles relative to the wavelength. Scattering does not affect S-band systems; it has a somewhat noticeable effect on X-band systems and it has considerable effect on K-band systems. At K-band frequencies, the scattering effect of rain and fog is so great that the equipment may be useless in bad weather. It is possible to see storms with X- and K-band radars because a sizable signal is returned to the radar by the scattering particles.

The choice of operating frequencies, discussed in chapter 1, is based on a compromise between improvement in fair-weather performance and increased vulnerability to adverse weather conditions, both of which accompany any increase in frequency.

Section III. EFFECT OF REFLECTED SIGNAL

21. Importance of Reflected Signal

The reflected signal is important only when it is picked up at the receiving point with a strength comparable to that of the direct signal. In most microwave installations, the reflected signal may be neglected either because it is absent or because it is weak as compared to the direct signal. When the reflected signal is present and strong, its importance lies in the fact that *it may either reinforce the direct signal or cancel it almost completely.* Whether the two signals tend to reinforce or cancel each other depends on their phase difference. As noted earlier (par. 15*b*), this phase difference results from the different path lengths of the direct and reflected signals. When the two signals are in phase with each other, or nearly so, they may be added directly; if they are of nearly equal strength, the combined signal is twice as strong as either one alone and it has a power level four times as great. But, when they are close to 180° out of phase with each other, *there will be partial cancelation, since the reflected signal (which is usually the weaker one) subtracts from the strength of the direct signal; when the two signal levels are equal or nearly so, cancelation may be substantially complete.* It is desirable, therefore, to consider a number of representative installations with a view to determining when the reflected signal must be considered and when it may be neglected at microwave frequencies. Among the determining considerations are—whether the transmission path is over land or water; whether the beam is compressed highly in height or width or both; whether the target (or receiving antenna) is moving or stationary; whether the signal is polarized horizontally or vertically; and the angle at which the signal strikes the reflecting surface. These considerations are discussed in the following paragraphs.

22. Overland and Overwater Operation

It has been pointed out that land is a rough surface and therefore a poor reflector at microwave frequencies, whereas water is a good reflector. A of figure 13, shows a point-to-point communications installation having an overland trans-

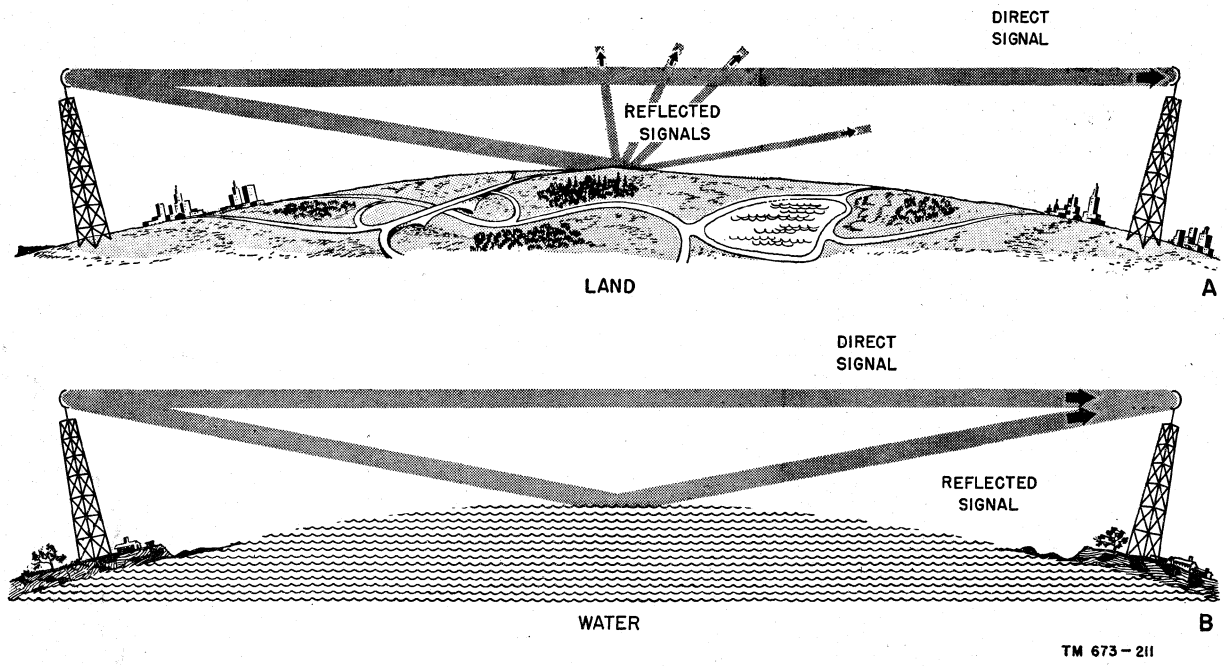


Figure 13. Reflected signals from land and water.

mission path and B shows a similar installation with an overwater transmission path. In A, the reflected signal is radiated in all directions from the point of reflection. Only a small portion of the energy reaches the receiving point, and the reflected signal may be neglected. In B, the water surface acts as a mirror, and the reflected signal is of nearly the same strength as the direct signal. In general, the reflected signal is unimportant and consequently may be neglected when the transmission path lies over land. Although the reflected signal is obviously important in the case of the overwater point-to-point communications system in B, it need not be considered necessarily in all overwater microwave installations.

23. Directivity of Transmitting Antenna

a. Tracking Radars.

(1) A tracking radar transmits and receives with an antenna having a highly directive pattern. The beam is compressed in both height and width, and is like a searchlight beam. Although it is always possible to draw the path of a reflected signal between the radar and the target, only a negligible amount of energy travels along that path.

(2) Figure 14 shows a tracking radar pointing at an aircraft. Because of the difference in the directions of the direct and the reflected paths, little or no reflected signal is transmitted to the target. The target, of course, returns echoes in all directions. In addition to the echo arriving along the direct path, an echo signal arrives at the radar by way of the reflected-signal path. This, however, is not picked up because the antenna is highly directional for both transmission and reception.

(3) If both the tracking radar and the target are close to the water surface, it is possible to have a situation like that shown in figure 13, where the direct and reflected paths are in nearly the same direction. Here, the reflected signal is important in the case of a point-to-point communication system.

(4) In the tracking radar, however, the target is moving rapidly and the direct and reflected signals cancel at one instant and add at another. The average effect is nearly the same as though the reflected signal were absent and, consequently, it may be neglected. It is important in the communications system only because both transmitter and receiver are stationary, so that the effect, whether cancelation or addition, is constant.

b. Search Radars.

(1) The antenna of a search radar has a pattern that is very sharp in the horizontal direction and fairly broad in the vertical plane (fig. 15). The result is that the reflected signal always must be considered in a shipborne search radar or a ground-based search radar that looks over water.

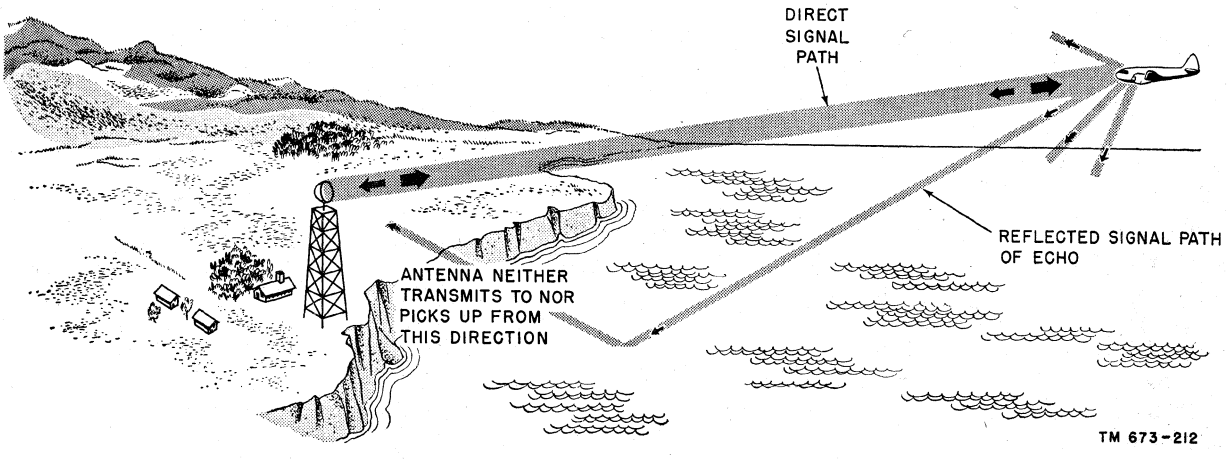


Figure 14. Direct and reflected signal paths for tracking radar.

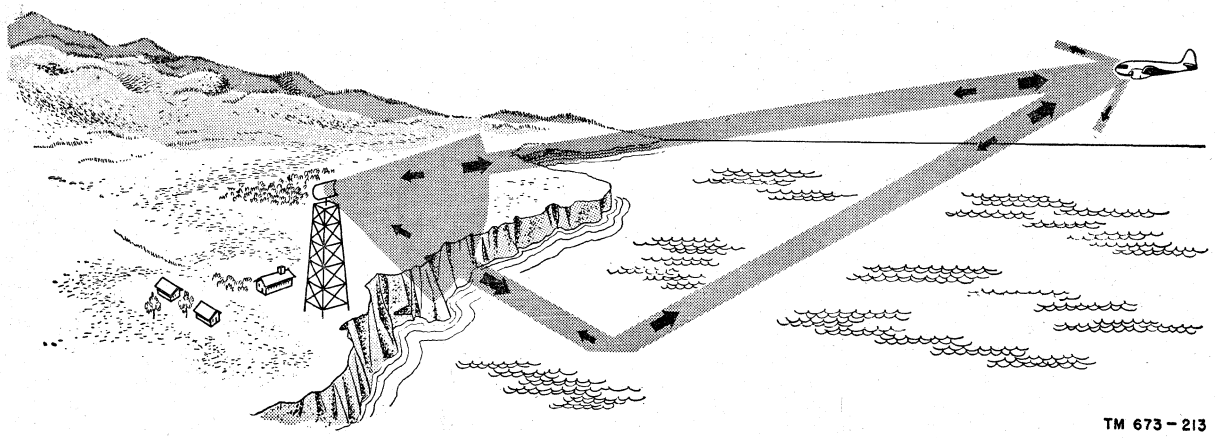


Figure 15. Direct and reflected signal paths for search radar.

(2) The reflected wave is present, as well, in an airborne search radar operated over water. The target, however, is relatively large—a ship or a coast line—so that cancelation of the direct and reflected signals at one point on the target is likely to be accompanied by addition of the signals at another point. The effect of the reflected signal, therefore, usually is neglected.

24. Horizontally and Vertically Polarized Waves

Briefly, land is a poor reflector and water is a good reflector at microwave frequencies. Land is rough at microwave frequencies, and it reflects signals in all directions in much the same way that a white painted wall reflects light; water is smooth, and it reflects signals in the same way as a mirror. (It is clear that a white painted wall reflects much of the light that falls on it, because a room with white walls is much lighter than one with dark brown or black walls. It is impossible however, to see an image, like the images observed in a mirror, by looking at a white painted wall.) There is, of course, a great deal of reflection from a land surface—every airborne radar operator has seen ground returns, or echoes from the ground. However, in the case of transmission over land between two points, the reflected signal is so weak compared to the direct signal that its effects may be disregarded. At microwave frequencies, it is necessary to consider the reflected signal only in an overwater point-to-point communications installation, a shipborne search radar installation, and a ground-based search radar installation which looks over water. The remainder of this section, therefore, is concerned only with the reflection of microwaves from a water surface.

a. It is possible to distinguish between two types of waves, called horizontally polarized waves and vertically polarized waves. The significance of this distinction to the present discussion lies in the different behavior of the two signals on reflection from a water surface.

b. A radio wave consists of both an electric and a magnetic field, each related to the other in a definite way. Both exist in the same region of space, and both move together at the same velocity in the direction of propagation (direction in which the wave is traveling). At any one point and any one instant, each field has a definite direction and strength. Along the direction of propagation, at any one instant, each field varies in strength and alternates in direction. A of figure 16, represents this variation.

c. Each solid arrow in A represents the direction and strength of the electric field throughout a surface called a wave front (par. 17*b*) perpendicular to the direction of propagation. The strength of the field throughout the wave front is proportional to the length of the arrow, and the direction of the field is the same as that of the arrow. Similarly, each broken arrow represents the direction and strength of the magnetic field through such a surface. B shows a small part of one wave front of the wave represented in A. The entire field contains an infinite number of these wave fronts, all moving in the same direction and at the same velocity.

d. As the wave fronts sweep past a fixed point

in the path of the wave, both the electric and the magnetic fields at that point alternate at the frequency of the signal, just as the voltage and current in an a-c circuit alternate at the generator frequency.

e. At distances greater than several wavelengths from the signal source, the two fields are in phase with each other; that is, the electric and the magnetic fields at a particular point reach maximum strength at the same instant. Also, at a particular instant, they both have maximum strength at the same point, as in A.

f. At every point in the path of the wave—that is, at every point on every wave front—and at every instant, the electric field and the magnetic field are perpendicular to each other, and both are perpendicular to the direction of propagation.

g. A *horizontally polarized wave* is one in which the direction of the electric field is horizontal, as in A. A *vertically polarized wave* is one in which the direction of the electric field lies in a vertical plane containing the path of the signal, as in C. It is not necessary that the electric field be vertical, but only that it lie in a vertical plane. Thus, if the signal path of the wave represented in C is inclined downward, the electric field is not vertical, but the wave nevertheless is polarized vertically.

h. As a matter of interest, the plane containing the direction of the electric field also contains the radiating element. Thus, a horizontal antenna radiates horizontally polarized signals and a vertical antenna radiates vertically polarized signals.

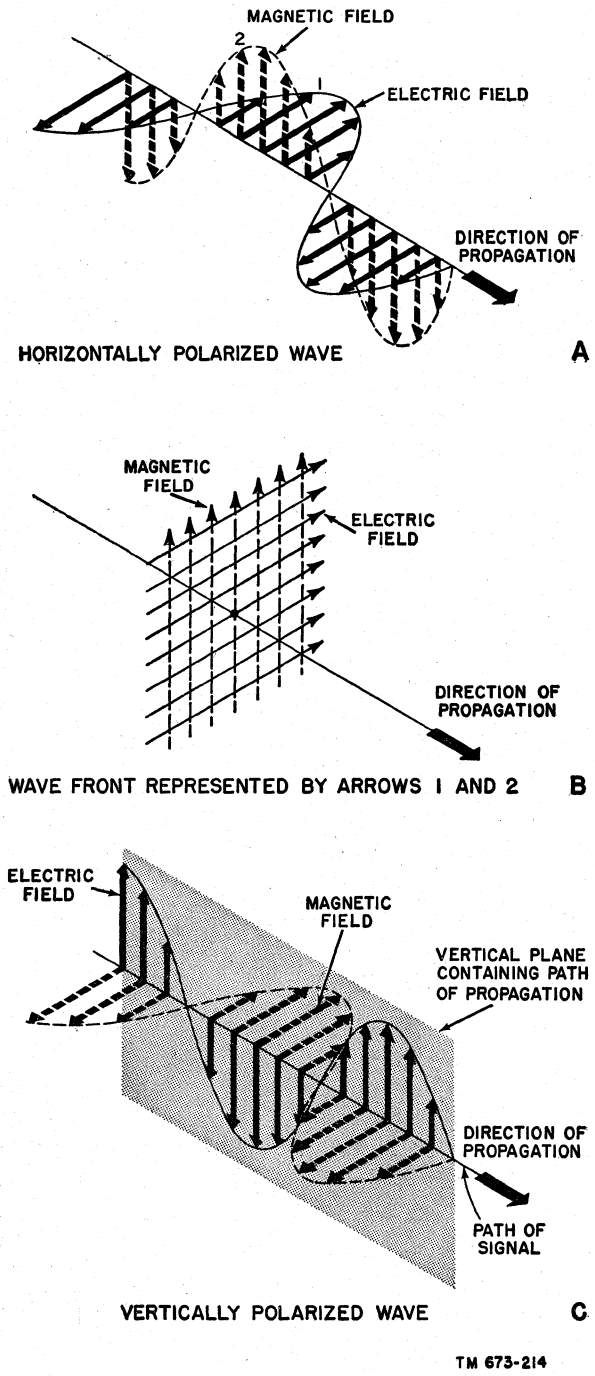


Figure 16. Horizontally and vertically polarized waves.

25. Reflection of Horizontally Polarized Signals

a. When a horizontally polarized wave strikes a horizontal surface, the electric field is parallel to the surface. This causes current to flow in the surface, and the current radiates the reflected signal. Each portion of the surface radiates reflected signals in all directions but, because of phase differences, the reflected signals from various portions of the reflecting surface cancel each other in all directions *except* the one taken by the reflected wave.

b. The incident wave (the wave falling on the surface) and the reflected wave differ in phase by 180° at the surface, so that one nearly cancels the other. The small electric field remaining after cancelation is responsible for the current flow in the surface. The greater the conductivity of the surface, the smaller is the electric field required to cause current flow, and the more nearly equal is the strength of the reflected wave to that of the incident wave. Sea water is a good conductor at microwave frequencies, and it is usual to say that the reflected signal has the same strength as the incident signal. A horizontally polarized signal

therefore is considered to be reflected from a water surface with a phase reversal (a phase shift of 180°) and with a negligible reduction in level.

26. Reflection of Vertically Polarized Signals

a. The reflection of vertically polarized signals from the surface of the sea is much more complicated than that of horizontally polarized signals. Both the phase and the level (strength) of the reflected signal change with the angle between the surface and the direction of the signal. In addition, the mechanism of reflection is more complicated than it is for horizontally polarized signals.

b. The direction of the incident signal may be specified by an angle called the *angle of incidence.* This is the angle between the path of the signal and a line perpendicular to the reflecting surface (fig. 17). When a vertically polarized signal strikes the surface of the sea at an angle of incidence close to 90°, the reflected signal is almost as strong as the incident signal. As the angle of incidence decreases, the reflected signal becomes weaker until it is only about one-fifth as strong as the incident signal at an angle of 84°; with further decrease in the angle, however, the reflected signal becomes stronger again.

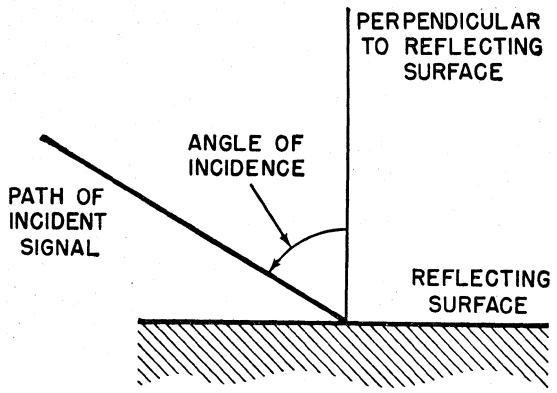


Figure 17. Angle of incidence.

c. When the angle of incidence is nearly 90°, the vertically polarized signal also suffers a phase reversal on reflection, but the phase change decreases to only 90° when the angle of incidence is 84° and continues to decrease as the angle of incidence is reduced further.

d. These figures apply to the S-band, but the same effects are present at both X- and K-bands, although the figures are different for each band. The phase angle or reflection of vertically polarized waves for a given angle of incidence increases with an increase in frequency. Consequently, microwave installations usually operate with horizontally polarized signals when the reflected signal is of importance.

27. Effect on Communications Installations

a. It already has been pointed out that the reflected signal may be neglected in an overland communications installation. The effect in an overwater installation is shown in figure 13.

b. When the direct path nearly touches the surface of the sea, the direct and reflected paths have the same length. Since the reflected signal suffers a phase reversal on reflection, it will arrive at the receiver precisely out of phase with the direct signal and at nearly the same level. The result will be almost complete cancelation, and reliable operation will be impossible.

c. If either antenna is made higher (or, of course, if both are made higher), the direct path clears the surface of the sea, and the length of the reflected path is increased. If the increase in height is chosen correctly, the length of the reflected path will be one half-wavelength longer than that of the direct path. This causes a phase difference of 180° between the direct and the reflected signals. The phase difference resulting from the difference in path lengths is added to the 180° phase difference which occurs because of the phase reversal on reflection, and the two signals will arrive in phase and add to each other. Additional increases in height cause further increases in the length of the reflected path, and this results in successive additions and cancelations of the signals. The calculation of the best antenna heights for a given installation is discussed in paragraphs 29–39.

28. Effect on Search Radar Installations

a. Path Difference.

(1) The effect of the reflected signal on the operation of a search radar is straightforward and can be understood easily by referring to figure 18. The radar antenna is located at A, and the direct path extends toward the target in direction AB. The reflected signal reaches the water at C and then is reflected in direction CD. Since the target is far distant, directions AB and CD are essentially the same. The effect is as though another radar were located at E, a point just as

far below the reflecting surface as the actual radar at A is above the surface. Distance AE then is equal to 2H, where H is the height of the radar antenna above the reflecting surfaces. Triangle AEC is an isosceles triangle, in which distance EC is equal to distance AC. Therefore, the length of the reflected path is not altered if it is assumed that the reflected signal originates at the image radar located at E.

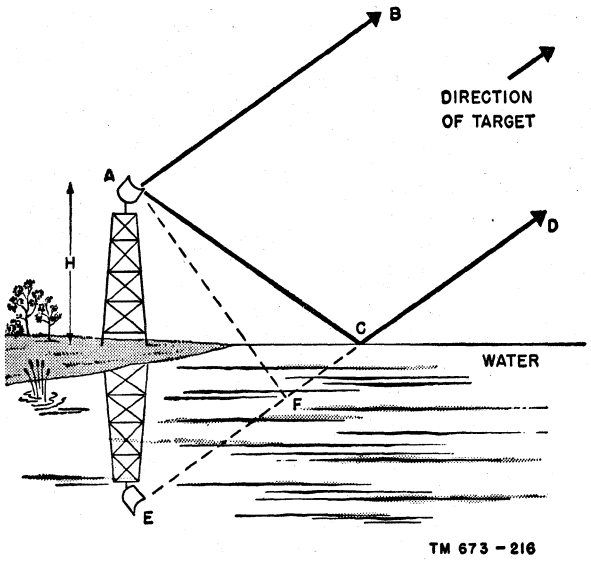


Figure 18. Path difference for search radar.

(2) Because of the phase reversal on reflection, however, the signals from the image radar must be precisely 180° out of phase with those from the real radar. The two signals at the aircraft are then the same as though they were received from the rear radar and its image. Next, AF is drawn perpendicular to path ED. Since AB and FD are parallel, the distance to the aircraft is the same, whether measured from A or from F. The geometrical difference in the two path lengths then is EF.

(3) Because the two transmitters are 180° out of phase, the direct and reflected signals will add at the target when EF is an odd number of half-wavelengths and will cancel when EF is an even number of half-wavelengths. There are, therefore, certain directions in which the radar cannot pick up a target except at extremely short ranges.

b. Vertical Polar Diagram.

(1) The combined radiation pattern of the actual radar and its image is the result of alternate addition and cancelation of the direct and reflected signals. It may be plotted on a height-range diagram to show the coverage of the radar; such a plot, called a vertical polar diagram, is shown in figure 19. (The construction of a height-range diagram of this sort is discussed in par. 36.)

(2) The parts of the pattern in which the radar has good performance (the directions in which the signals add) are called *lobes*, and the parts in which the radar cannot pick up targets are called *gaps*. The VPD (vertical polar diagram) (fig. 19) shows, for a radar located 1,000 feet above sea level, gaps extending from 70 to 90 miles at 15,000 feet, from 100 to 135 miles at 10,000 feet, and so on. There is always a gap at the horizon, because both path lengths are the same in that direction.

(3) As the wavelength at which the system operates is decreased, the lobes and gaps become narrower and closer together. This results from the fact that a smaller change in direction is required for a given change in the phase difference between the signals at the target (fig. 18). A given change in direction makes a given change in distance EF. Such a change, however, may be only a fraction of a wavelength at very-high frequencies or several wavelengths at microwave frequencies. In practice, the separation of the radar and its image, as measured in wavelengths, is the determining factor. If distance AE is several hundred wavelengths, a small change in direction is still able to change EF by several wavelengths.

(4) The VPD of a vhf radar shows relatively wide gaps, because the height of the radar may be as little as 8 or 10 wavelengths (the height is one-half of AE). At microwave frequencies, however, the height of the radar is at least 100 wavelengths, and may be several thousand wavelengths. The result is that the

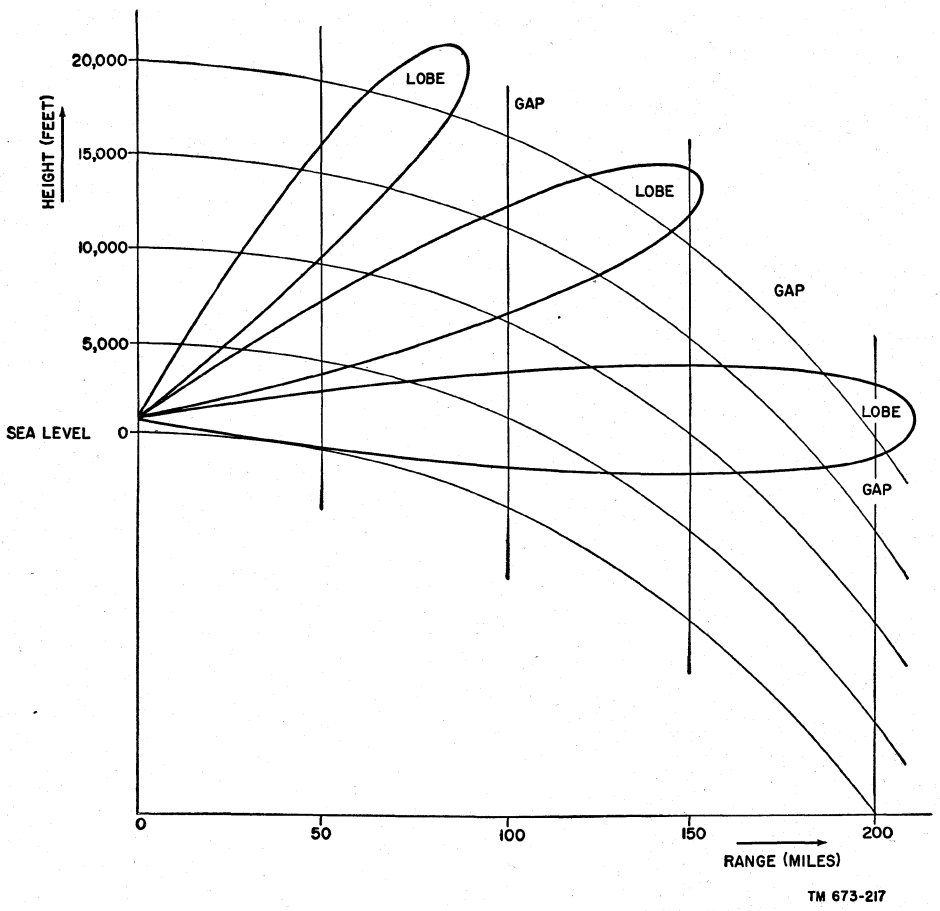


Figure 19. Vertical polar diagram.

gaps are so narrow and the lobes are so closely spaced on the VPD that the general effect is one of uniform vertical coverage. In most cases, therefore, the effect of the reflected signal is simply to increase the apparent power of the search radar. It must be emphasized, however, that this is *not* true of vhf radars.

Section IV. LINE-OF-SIGHT LIMITATIONS

29. Radio Line of Sight

a. The path followed by a light wave in traveling from one point to another is called the *optical line of sight* between the two points. The path followed by a microwave or vhf signal under standard atmospheric conditions is called the *radio line of sight* or, less frequently, the microwave or vhf line of sight between the two points. For most purposes, optical lines of sight may be considered as straight lines. Radio lines of sight, except when they are vertical, curve downward slightly (par. 19*b*).

b. It has been explained that the effects of the reflected signal can be neglected at microwave frequencies except in an overwater point-to-point communications system. (This is true even in the case of an overwater search radar, where the level of the reflected signal ordinarily is comparable to that of the direct signal.) In an overwater point-to-point communications system, successful operation requires the existence of a suitable path for the direct signal *and* suitable phase relations between the direct and reflected signals at the receiving point. In all other cases, successful operation requires nothing more than a suitable path for the direct signal. (It is assumed, of

course, that suitable equipment with adequate power output for the installation is available.)

c. Under conditions of standard propagation, the direct wave follows a radio line of sight. Usually, therefore, it is necessary only to determine whether an unobstructed radio line of sight connects the two points between which microwave operation is desired.

30. Overland Point-to-Point Communications System

As stated in the preceding paragraph, successful operation of a microwave overland point-to-point communications system (assuming adequate equipment) under conditions of standard propagation requires nothing more than the existence of an unobstructed radio line of sight connecting the two points between which operation is desired. Two ways of determining whether such a line of sight exists are by visual inspection of the path, or by construction of a profile diagram.

31. Inspection of Path

In many cases, particularly where the two points are not too far apart, it is possible to make a direct inspection of the path. If the proposed position of one antenna can be seen (optically) from the proposed position of the other antenna, satisfactory operation usually will be possible. The optical line of sight joining the two antenna positions is a straight line, whereas radio line of sight curves downward. Consequently, the radio line of sight is *above* the optical line of sight at all points except the antenna positions, where the two lines of sight coincide (fig. 20). If, therefore, there are no obstructions along the optical line of sight, and no obstructions above it, the radio line of sight will be clear. It is usual to insist that there be no obstructions within 50 feet or so of the optical line of sight.

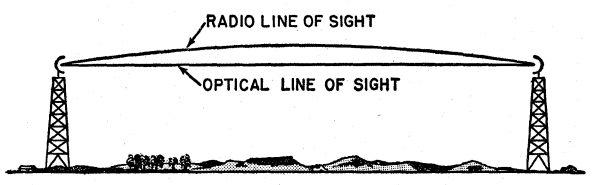


Figure 20. Optical and radio lines of sight.

32. Profile Diagrams

a. When it is impossible to inspect the path between two points, or when the results of an inspection are not clear, it is necessary to make a *profile diagram* of the path in order to decide whether there is an unobstructed radio line of sight for the direct signal. A profile diagram is nothing more than a picture that shows the height of the land along a path between two points.

b. Consider, for example, two antenna towers, each 100 feet high, separated by 20 miles of smooth, flat country. A, of figure 21, shows a profile diagram based on a flat earth; that is, the earth is *assumed* to be flat. The size of the diagram, of course, has been reduced; originally, it was drawn with a horizontal scale of 2 miles per inch and a vertical scale of 100 feet per inch. The reason for the different horizontal and vertical scales will be discussed later. Since the diagram is based on a flat earth, the surface of the earth is indicated by a straight line.

c. B shows the same antennas but takes account of the curvature of the earth. Referring to appendix I, it is seen that the surface of the earth between two points rises a distance equal to $S^2/6$ feet above a straight line (through the earth) between the points, where S is the distance in statute miles (of 5,280 feet) between the points. For the present instance, S is 20 miles, and $S^2/6$ is 67 feet. The surface of the earth is shown by an arc of a circle, and the arc is chosen so that its midpoint is 67 feet above the straight line between the tower bases. The distance of 67 feet, like the heights of the antenna towers, is measured on the scale of 100 feet per inch.

d. Since this profile diagram, in B, takes into account the curvature of the earth, any *optical* line of sight is a straight line. The radio line of sight between the antennas curves downward, and therefore lies above the optical line of sight. The maximum distance between the two lines of sight is shown in the diagram as 17 feet, and may be found by the following method.

e. For working with microwaves, it is desirable to use a profile diagram based on a *modified earth.* On such a diagram, the *radio* line of sight, not the optical line of sight, appears as a straight line. The modified earth has a radius that is four-thirds of the radius of the real earth (4,000 miles) and its surface therefore is curved somewhat less. It is shown in appendix I that the surface of the modified earth rises only $S^2/8$ feet above a straight line

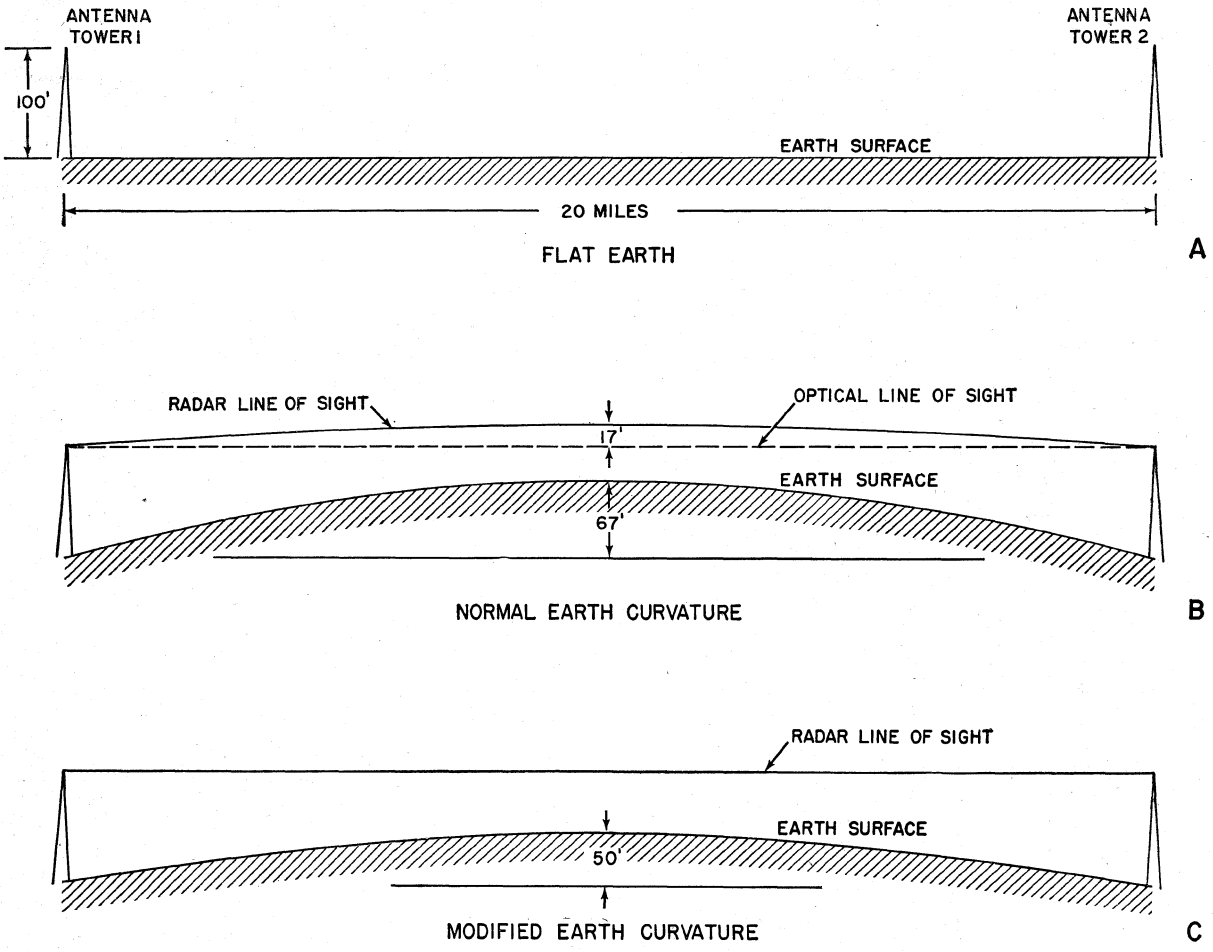


Figure 21. Radio line of sight over flat, normal and modified earth.

between two points separated by a distance, S. C shows a profile diagram based on the modified earth. As before, the surface of the earth is shown by an arc of a circle, but the midpoint of the arc is now only $S^2/8$ feet, or 50 feet, above the line joining the bases of the antenna towers. This distance differs from the corresponding distance of 67 feet on the normal earth diagram shown in B by 17 feet, the true maximum distance between the radio and the optical lines of sight.

f. The construction of the modified earth profile diagram in C is equivalent to a distortion of the diagram in B, which makes the downward-curving earth surface a little less curved—by precisely the right amount to distort the curved radio line of sight into a straight line. Therefore, as mentioned in paragraph 18*b*, all radio lines of sight *drawn on a modified earth profile diagram*, in C, are drawn as straight lines.

33. Use of Profile Diagram

a. As an example of the use of a profile diagram, consider the following problem: Two relay units are to be installed as part of a microwave point-to-point communications system. Each relay unit is designed to receive signals and pass them on, so that reliable operation between the two units will be necessary if the communications system is to serve its purpose. Proposed sites have been chosen for the units, as shown on the map in figure 22. One relay unit is to be located on top of a rocky ridge 1,900 feet above sea level, and the other is to be located on a similar ridge 1,100 feet above sea level. The country between the two locations is low except for a wooded ridge 800 feet above sea level, 50 miles from the higher location. Assuming that the antenna towers are to be 100 feet high, the antennas will be 2,000 feet

and 1,200 feet above sea level. The problem is this—Can the two relay units operate successfully across the wooded ridge between them, or will a third relay unit be required?

b. The problem is solved easily by referring to the profile diagram in figure 23, which is constructed as follows: Lay out horizontal line AB and vertical line CD, making sure that they are perpendicular to each other. Using a suitable horizontal scale (20 miles to the inch in this diagram), lay out the positions of the antenna locations and the wooded ridge which is under consideration as a possible obstruction. All horizontal distances are measured from the center of the path. The site of the higher antenna is therefore 2 inches to the left of E (the intersection of the horizontal and the vertical lines) or 40 miles from the center of the path, at F. The lower antenna is at G, an equal distance on the other side of E. The wooded ridge is at H, ½ inch, or 10 miles, from E.

c. Since *S*, the distance between the two antennas, is 80 miles, the height of the midpoint of the arc is 800 feet (using value $S^2/8$ which applies to the modified earth). Choosing a convenient vertical scale, say 800 feet to the inch, locate I, which is 1 inch, or 800 feet, above E. Next, locate J, which is the center of a circle passing through F, G, and I, and draw the arc through these points. J is located by trial and error. Place the point of a compass somewhere on the vertical line below E, set the radius so that

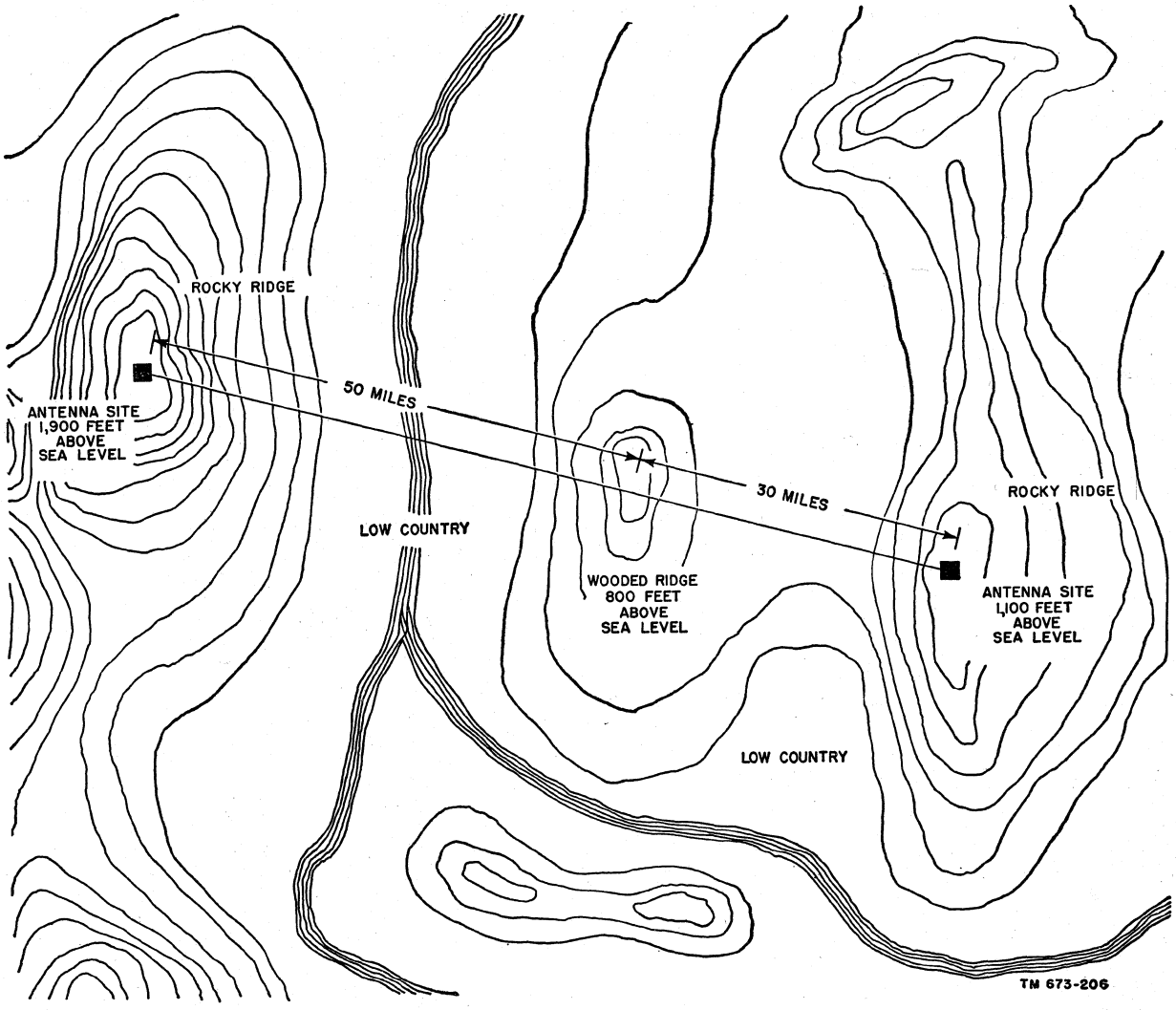


Figure 22. Map of profile diagram problem.

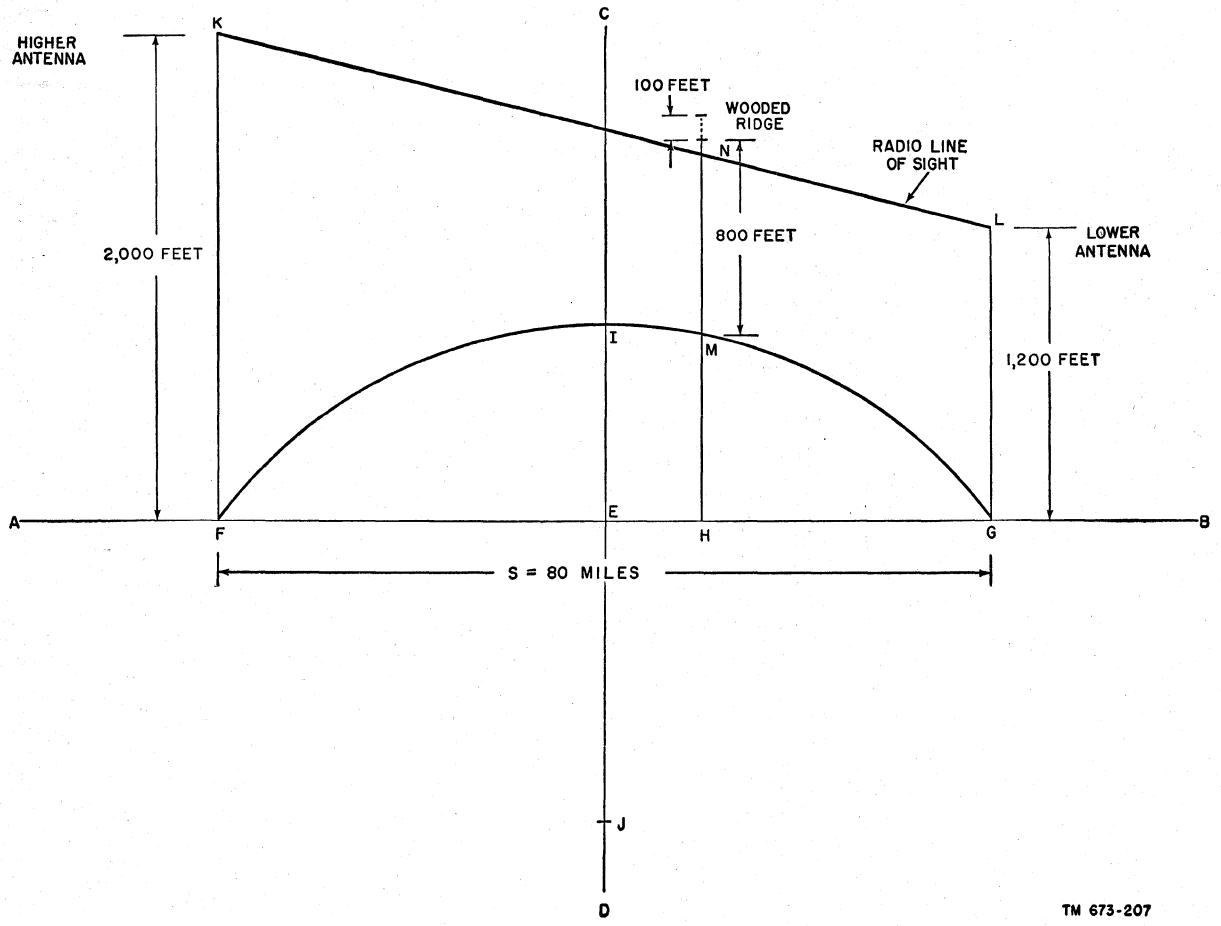


Figure 23. Profile diagram of problem.

the circle will pass through I, and see whether the same circle passes through F and G. If it falls outside these points, select a new center higher up, set a new and shorter radius, and try again. If, finally, a circle is obtained which passes through F and I but not through G, or through G and I but not through F, the original lines, AB and CD, are not perpendicular and it will be necessary to start again.

d. Having drawn the arc, locate the higher antenna at K by measuring upward 2.5 inches, or 2,000 feet, from F. Note that this measurement is made in a direction perpendicular to AB, *not* in a direction perpendicular to the arc. Locate the lower antenna at L by measuring upward 1.5 inches, or 1,200 feet, from G. The top of the wooded ridge, N, now is drawn in. It is located directly above H, and its height is 1 inch, or 800 feet, above M on the arc that represents the surface of the earth. The line upward from H may be extended an additional 100 feet to allow for the height of the trees. This has been done on the drawing, and the extension is shown as a dotted line. Finally, draw the radio line of sight as a straight line between the two antennas, K and L. It is clear from this profile diagram that the two relay stations will not operate properly unless a third unit is located on top of the intervening ridge.

34. Error in Profile Diagram

a. The profile diagram shown in C of figure 21, is based on a horizontal scale of 20 miles to the inch, or 105,600 feet to the inch. It is not practical to use the same scale for vertical distances, since the midpoint of the arc is only 800 feet above the base line, and this would be less than 1/100 inch on the diagram. When the vertical scale is expanded—in this diagram, 800 feet per inch—some error is introduced. The error is a height error; that is, the arc that represents the

surface of the earth is not exactly where it should be. The error is zero at the center of the arc and at each end, and reaches its greatest value at two points .15(S) from each end. In the problem example given in paragraph 33, S is 80 miles, and the maximum error occurs at points 12 miles from each antenna. (The distance of 12 miles is measured from the end of the arc toward the center along the straight line.)

b. The maximum error in feet is $(S^2/8)\ (H/W)^2$, where S is the distance in miles between the points on which the profile diagram is based, H is the height of the midpoint of the arc in inches on the diagram, and W is the width of the arc in inches on the diagram. In the example above, S is 80 miles, so that $(S^2/8)$ is 800 feet. H is 1 inch (800 feet on a scale of 800 feet per inch), and W is 4 inches (80 miles on a scale of 20 miles per inch). Quantity (H/W) then is equal to ¼ and $(H/W)^2$ is 1/16. The maximum error, therefore, is 800/16, or 50 feet. The error at the point of interest is, of course, smaller, since the maximum error occurs at points only 12 miles from the antennas. If it is necessary to reduce the error, a different vertical or horizontal scale can be used, so that H/W becomes smaller.

35. Radar Coverage

In considering a proposed point-to-point communications installation, it is necessary only to predict whether operation will be possible between the proposed sites. In a radar installation, however, it is necessary to predict the coverage available from a proposed site—that is, the ranges at which various targets may be detected. In the vhf range of frequencies, coverage may be affected by gaps (par. 28*b*), although there are a number of methods of *filling* the gaps. At microwave frequencies, the vertical polar diagram is effectively free from gaps, and with adequate equipment the coverage is limited only by obstructions which interfere with the radio line of sight to the target.

36. Height-Range Diagrams

Radar coverage usually is shown on a height-range diagram such as that in figure 19. Height-range diagrams are similar to profile diagrams, and they are based on the same mathematics. The only important difference is that the equipment site is shown at the midpoint of the arc on a height-range diagram, because the maximum error in the diagram occurs near the longest ranges instead of near the shortest ones. Since coverage often is different in different directions, a separate height-range diagram must be prepared for each direction that is of particular interest. A height-range diagram usually shows coverage in a single direction and is similar to half of a profile diagram.

a. Referring to figure 24, a height-range diagram may be constructed in the following steps:

(1) Draw horizontal line AB and vertical line CD, making sure that they are perpendicular to each other. Although any convenient distance may be taken as the basis for a height-range diagram, 200 miles is probably a usual choice. In figure 24, a horizontal scale of 50 miles to the inch was used, and F is located 4 inches, or 200 miles, to the right of E, the intersection of lines AB and CD.

(2) Since EF, which is 200 miles, is only half of the base line of the equivalent profile diagram, the height of the arc is calculated using 400 miles for S, so that $S^2/8$ is 20,000 feet. A vertical scale of 10,000 feet to the inch has been used in figure 24, and the next step is the location of G, which is 2 inches, or 20,000 feet, above E.

(3) Next, find H (on CD) which is the center of a circle passing through F and G, and draw in arc GF, which represents the surface of the earth. I, J, K, and L are located at ½-inch (5,000-foot) intervals above H, and these intervals are used as centers for the 5,000-, 10,000-, 15,000-, and 20,000-foot arcs that all are drawn with the same radius used for arc GF.

(4) Finally, M, N, and O are located at intervals of ½ inch, or 50 miles, and the vertical range lines are drawn in parallel to CD.

b. The maximum height error occurs at a range .15 (S) to the left of point F. Since S is 400 miles, .15 (S) is 60 miles, and the greatest error is at a range of 140 miles. The size of the maximum error is $(S^2/8)\ (H/W)^2$, where S, H, and W have the values (par. 34*b*) associated with the equivalent profile diagram. Thus, S is 400, and $(S^2/8)$ is 20,000 feet. H is 2 inches, W is 8 inches (twice the width of the height-range diagram), and, therefore, $(H/W)^2$ is 1/16. Consequently, the maximum height error, which occurs at a range of 140 miles, is 1,250 feet. The error drops to zero at ranges of zero and 200 miles (the midpoint and end of the arc). If a vertical scale of 20,000 feet to the

252020°—53——3

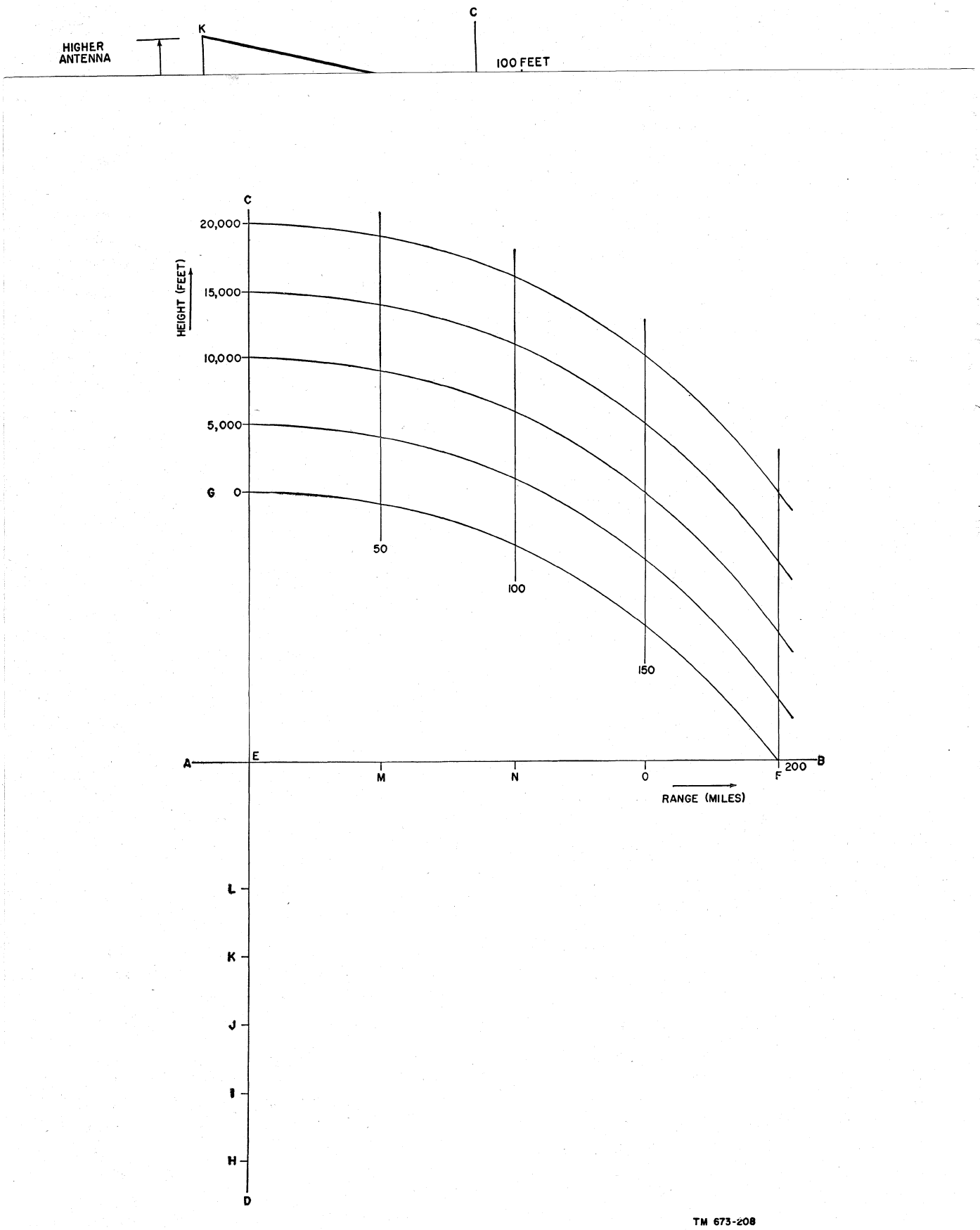


Figure 24. Height-range diagram.

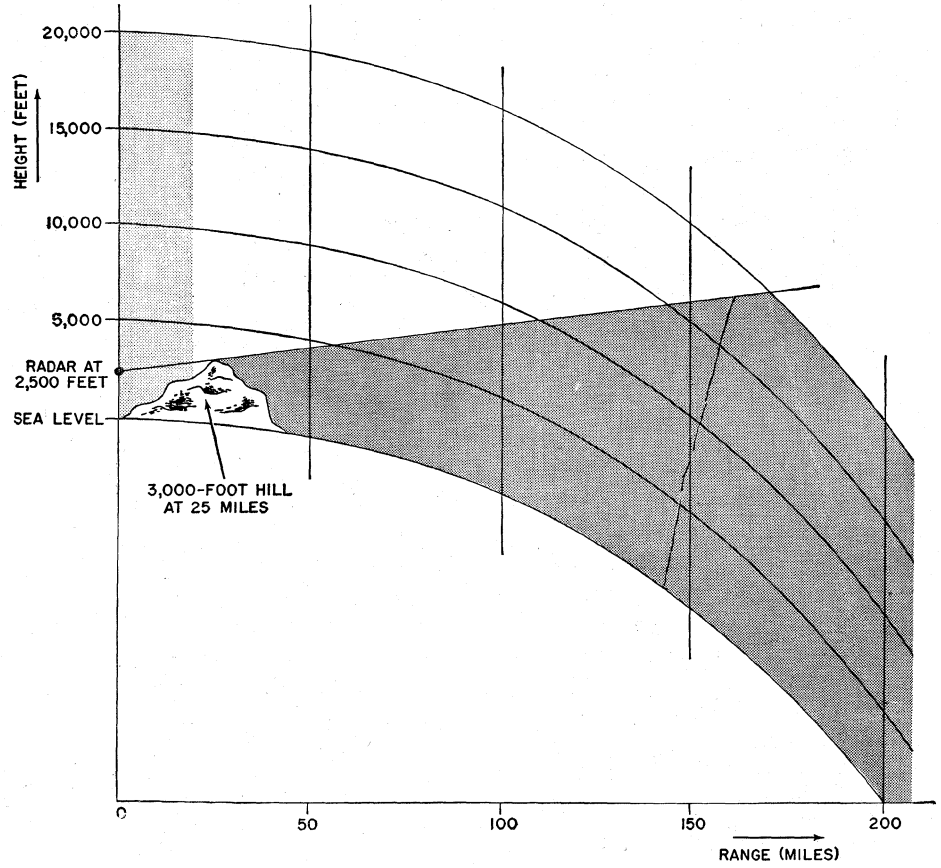
inch is used (or if the horizontal scale is changed to 25 miles to the inch), the maximum error will drop to 312.5 feet.

37. Radar Coverage Over Land

Figure 25 illustrates the method of showing radar coverage on a height-range diagram. This diagram is made for a radar sited 2,500 feet above sea level. The only significant obstruction in the direction for which the diagram is made is a hill 3,000 feet high at a range of 25 miles. The profile of the hill is plotted in the manner described in paragraph 33*d*, height being measured vertically upward from the arc representing the surface of the earth. A radio line of sight is drawn as a straight line on the diagram, from the transmitter through the crest of the hill and beyond. Below this line, in the shaded region, there is no coverage under standard conditions because of obstruction by the hill. The shaded region at the left, at ranges shorter than 20 miles, is a region of poor coverage. Echoes from targets in this region may be obscured by echoes from the ground. The clear region, above the limiting radio line of sight and at ranges greater than 20 miles, is the region of good coverage.

38. Radar Coverage Over Water

A much simpler problem is posed when most of the line of sight lies over water. A height-range



MINIMUM RANGE	20	MILES					
MAXIMUM RANGE	57	MILES	ON	TARGET	AT	5,000	FEET
	110	"	"	"	"	10,000	"
	143	"	"	"	"	15,000	"
	170	"	"	"	"	20,000	"

TM 673-209

Figure 25. Radar coverage.

diagram seldom is used, because the maximum range on a target at a given altitude can be calculated easily.

a. The geometry of the problem is shown in A of figure 26. The maximum range, S, is the sum of S_1 and S_2, where S_1 is the distance from the radar to the horizon, and S_2 is the distance from the horizon to the target. In order to find S, then, it is necessary only to know the law that determines the distance to the horizon when the height of a point is known. The problem is reduced to the solution of the right triangle, shown in B. By Pythagoras' theorem,

$$(R+H_1)^2=R^2+S_1^2$$

$$R^2+2RH_1+H_1^2=R^2+S_1^2$$

$$2RH_1+H_1^2=S_1^2$$

$$S_1^2=H_1(2R+H_1)$$

Since H_1 (the antenna height) is negligibly small compared to $2R$ (twice the earth's radius), the second factor on the right $(2R+H_1)$ is nearly equal to $2R$. Therefore, $(2R+H_1)$ may be replaced by the approximation $2R$, so that

$$S_1^2=2RH_1$$

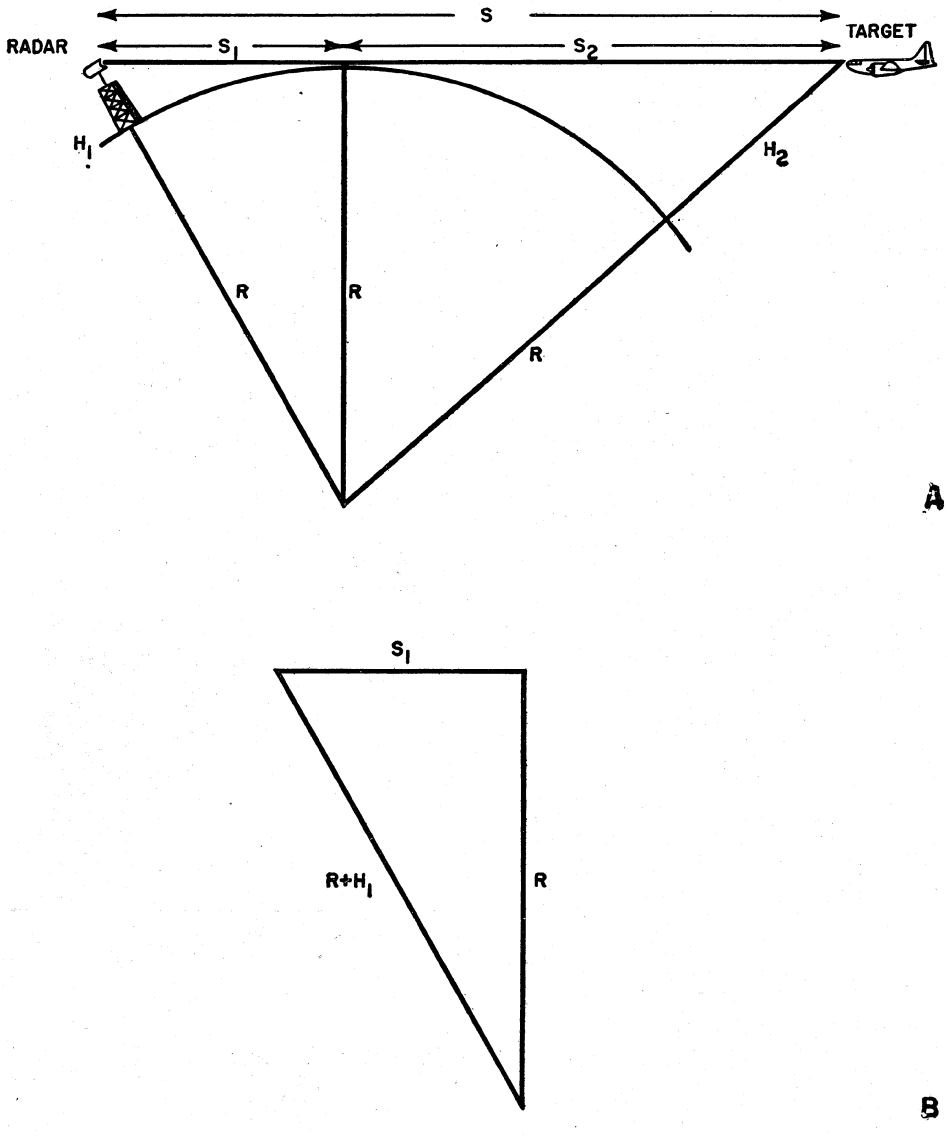


Figure 26. Distance to horizon.

This equation is satisfactory if S_1, R, and H_1 are in the same units. However, S_1 and R generally are expressed in miles, whereas H_1 is in feet. H_1 must be reduced to miles by the factor 1/5,280, since a foot is equal to 1/5,280 mile, and thus the equation becomes

$$S_1^2=2RH_1/5{,}280=RH_1/2{,}640.$$

b. To determine the distance to the *optical* horizon, R is set equal to 4,000 miles and

$$S_1^2=H_1(4{,}000/2{,}640)=H_1(500/330)=H_1(50/33)$$
$$S_1^2=H_1(100/66)$$

The quantity, 100/66, is nearly equal to 3/2, or 1.5, so that, by approximation,

$$S_1^2=1.5H_1$$

c. To determine the distance to the *radar* horizon, R is replaced by 4/3 times 4,000 miles, and 1.5 must be replaced by 4/3 times 1.5, or 2.0. The equation then becomes

$$S_1^2=2H_1$$

For a point at a height H above the water, then, the optical horizon is at a distance S_1 equal to $\sqrt{1.5H_1}$ or $1.23\sqrt{H_1}$, and the distance to the radar horizon is $\sqrt{2H_1}$, or $1.41\sqrt{H_1}$. Similarly, the radar distance, S_2, is $\sqrt{2H_2}$, or $1.41\sqrt{H_2}$. The range of the radar, finally, is

$$S=S_1+S_2=\sqrt{2H_1}+\sqrt{2H_2}$$

d. Suppose, for example, that the radar is set at a height of 50 feet, and the target aircraft is at an altitude of 5,000 feet (A of fig. 26). The range from the radar to the horizon is S_1; this is calculated by taking the square root of twice the antenna height, or $\sqrt{2H_1}$. The range from the horizon to the target is S_2; this is calculated by taking the square root of twice the altitude of the target, or $\sqrt{2H_2}$, and where $2H_1$ equals 100 feet and $2H_2$ is equal to 10,000 feet. The maximum possible range will be

$$\sqrt{100}+\sqrt{10{,}000}=10+100=110 \text{ miles}$$

39. Overwater Point-to-Point Communications Systems

In an overwater point-to-point communications system, the reflected signal is important. It is necessary, therefore, in choosing antenna heights for a proposed installation, to make sure that the direct and reflected signals do not cancel each other at the receiving point. It is possible, of course, to calculate the best antenna heights for use under standard conditions, but the computation is tedious and the results will not be correct for nonstandard conditions. Under the circumstances, antenna heights may be chosen by the rule-of-thumb method described here. Choose tentative antenna heights H_1 and H_2 which satisfy the relation,

$$\sqrt{2H_1}+\sqrt{2H_2}=S,$$

where H_1 and H_2 are in feet, and S is the distance between the proposed sites in miles. Next, calculate a correction height, h, where

$$h=660\sqrt{S/f}.$$

S is defined above, f is the frequency of operation in mc per second, and h is in feet. The required height for each antenna is, then, the sum of the tentative height and the correction height, so that, by this rule, the antenna heights should be (H_1+h) and (H_2+h). If these values are inconvenient, choose new tentative heights and make a new computation.

Section V. OTHER LIMITATIONS ON MAXIMUM RANGE

40. Maximum Ranges

In general, communication between two points is impossible when no path exists for a direct signal. This statement now can be phrased as follows: Communication between two points is impossible unless an unobstructed radio line of sight exists. This does not mean, however, that communication is automatically possible whenever the line of sight is clear. Even with a clear line of sight, the signal level may be too low to be useful because of insufficient transmitter power, inadequate receiver sensitivity, or poor reflection from the target (in the case of a radar). Besides this, there is always the possibility that the direct and reflected signals will arrive at the receiving point with nearly the same amplitude and with a phase difference of 180°. In this event, they will cancel each other, and communication will be impossible. Finally,

various atmospheric effects may cause the curvature of the path followed by the signals to vary or may attenuate the signal severely so that communication is not possible.

41. Maximum Range of Point-to-Point Communications System

The maximum range over which a point-to-point communications system can operate reliably usually is limited by line-of-sight considerations. That is, the problems of long-range operation usually reduce to a question of finding satisfactory transmitting and receiving sites and building antenna towers that are sufficiently tall. With lightweight or portable equipment, however, it happens occasionally that the maximum reliable operating range is limited by considerations of available signal power. It is difficult to compute the maximum range over which a given system will operate reliably, but it can be shown that the maximum range depends on the characteristics of the system.

a. Power Density at Receiving Site. For any type of equipment, the power delivered to the receiver is proportional to the power picked up by the receiving antenna. This, in turn, is proportional to the area of the receiving antenna and to the power density (power per unit area in a plane perpendicular to the direction of propagation) at the receiving site. The effectiveness of the transmitting equipment, therefore, is measured in terms of power density delivered at the receiving site, and this is proportional to P_tA/R^2, where P_t is the power output of the transmitter, A is the area of the transmitting antenna, and R is the distance from the transmitting site to the receiving site. P_t is entered in the expression as the first power in the numerator, since doubling the transmitter power will double the power density. A is equally logical. For example, if the width of the transmitting antenna is doubled, the width of the radiated beam will be halved. Therefore, the power in the beam will be distributed over one-half of the original area. This, of course, doubles the power density, and doubling the antenna width doubles the area of the antenna. The explanation of f is a little more complex. Doubling the frequency halves the wavelength and allows the antenna dimensions to be halved without altering the radiation pattern. If, therefore, the frequency is doubled, the antenna area may be cut to one-fourth of its original value (by halving both of its dimensions) without altering the power density. Following this, the antenna area may be quadrupled by restoring the antenna to its original size, and this will quadruple the power density. The power density thus is quadrupled by doubling the frequency, if the antenna area is unchanged. Finally, R is in the denominator as the second power, since doubling the range to the receiving point quadruples the area covered by the beam at the receiving point and cuts the power density to one-fourth. As stated before, the power density delivered at the receiving site is proportional to P_tA/R^2.

b. Power Delivered to Receiver. The power delivered to the receiver for a given power density is proportional to the area of the receiving antenna, since the larger the antenna the greater the amount of power intercepted. Therefore, since the power density delivered at the receiver site is proportional to P_tA/R^2, the power delivered to the receiver is proportional to P_tAaf^2/R^2, where a is the area of the receiving antenna.

c. Maximum Operating Range.

(1) If the power delivered to the receiver is termed p, the expression above may be stated thus:
p is proportional to P_tAaf^2/R^2
The minimum usable receiver power, p', will be delivered to the receiver at the maximum reliable operating range, R', where
p' is proportional to P_tAaf^2/R'^2
and, by inversion,
R'^2 is proportional to f^2AaP/p'
From this, it follows that
R' is proportional to $f\sqrt{AaP/p'}$

(2) On the basis of these equations, the maximum reliable operating range theoretically can be doubled by doubling the frequency. Practically, this is not true, because bad weather affects high-frequency signals more than it affects low-frequency signals. This offsets the increase in maximum range to some extent, and even may make matters worse if the frequency is raised too far. In relation to other quantities, the expression is more reliable. To double the maximum range by a change in a single quantity, it is necessary to quadruple the area of one of the antennas, making either A or a four times as large as before; to quadruple the power output

of the transmitter, making P four times as large as before; or to quadruple the power sensitivity of the receiver, making p' only one-fourth as large as before. The improvement, however, will be offset slightly by the fact that atmospheric effects will be more pronounced over the longer path when the range is increased.

42. Maximum Range of Radar System

The maximum operating range of a radar system normally is limited also by line-of-sight considerations. As with communications systems, however, lightweight or portable equipment may have its maximum range limited by considerations of signal power. The dependence of the maximum range on the system characteristics is easy to derive, since it follows directly from the expressions found in the preceding paragraph. The power density at the target is, as before, proportional to P_tA/R^2. This could be multiplied by the target area to find the power falling on the target. Instead, it is multiplied by a quantity called the *effective target area* in order to obtain the power in the portion of the reflected signal that travels back toward the radar. If the effective target area is b, the target acts like a transmitter with a power density proportional to P_tAbf^2/R^2. This expression then is divided by R^2 to obtain P_tAbf^2/R^4, which is proportional to the echo power density at the radar. Finally, it is multiplied by A, the area of the radar antenna, to obtain $P_tA^2bf^2/R^4$, which is proportional to the echo power picked up by the radar antenna. Then, as before,

p is proportional to $P_tA^2bf^2/R^4$,
p' is proportional to $P_tA^2bf^2/R'^4$,
R'^4 is proportional to f^2A^2bP/p',

and the maximum range of the radar is proportional to $\sqrt[4]{f^2A^2bP/p'}$. It is important to note in this expression that the maximum range varies as the fourth root of P/p'. Because of this, a 10-percent change in the power output of the transmitter in the power requirement of the receiver will make only a 2.5-percent change in the maximum range on targets of a given type.

43. Summary

a. At low and medium frequencies, communication usually is possible between two points if a sufficiently powerful transmitter and an adequately sensitive receiver are available. At microwave frequencies, communication is impossible unless the microwave signals also have an unobstructed path, because the sky wave signal is not present at these frequencies. The ground wave signal occasionally is present, but it is of no practical importance. Communication between two points therefore is dependent on the direct signal.

b. The path followed by a microwave signal curves slightly downward under normal atmospheric conditions. Such a path is called a radio line of sight, as distinguished from an optical line of sight, which is a straight line. Communication between two points is possible only when there is an unobstructed radio line of sight joining one point to the other.

c. Under certain atmospheric conditions, the curvature of the radio line of sight is altered. Such conditions are called anomalous conditions, and propagation thereunder (along paths having nonstandard curvature) is called anomalous propagation. The usual effect of anomalous propagation is to increase the maximum possible range of microwave operation by making it possible for signals to reach points beyond the horizon. This results from an increase in the curvature of the radio line of sight. Occasionally, the effect is in the opposite direction. The curvature is reduced, and the radio line of sight even may curve upward. Microwave operation under such conditions is nearly impossible, but such conditions are rare.

d. X-band and K-band signals are scattered by rain or snow, and long-range radar equipment normally operates in the S-band to avoid this difficulty. Short-range equipment, however, normally operates in the X-band in order to make use of the greater directivity available at the shorter wavelengths from the antennas of reasonable size.

e. Microwave communication is primarily by means of the direct signal. The reflected signal is present at an appreciable level in overwater operation, but may be neglected except in the case of overwater point-to-point communications systems.

f. The performance of a point-to-point communications system between two given points may be predicted by constructing a profile diagram of the path. This makes use of a distorted earth, having four-thirds the radius of the actual earth. A radio line of sight, when plotted on such a

diagram, appears as a straight line. A similar diagram, called a height-range diagram, is used for predicting the coverage of a radar at a given site. For predicting overwater operation, it is more convenient to make use of formulas that are given in the text.

g. In most cases, the maximum range of microwave equipment is limited by line-of-sight considerations. With portable equipment, however, transmitter power is occasionally the limiting factor. Given an unobstructed line of sight, the maximum range for point-to-point operation is proportional to the square root of the transmitter power and the maximum range for radar operation is proportional to the fourth root of the transmitter power.

44. Review Questions

a. Reliable operation over distances of 100 miles or less generally is possible with radio systems operating below 30 mc per second. Is this the case with microwave systems?

b. Name several factors affecting the microwave propagation characteristics of a path between two points.

c. Name the four signals involved in the transmission of energy at radio frequencies from one point to another. Need all four signals be present?

d. Which signal is of primary importance to microwave installations? What is the path it takes?

e. In what way will energy arriving along a reflected-signal path over water differ from a similar path over land? Why?

f. Assume a strong direct signal and a strong reflected signal between two points. Is the combined signal level at the receiving end necessarily high?

g. When the phase difference between direct and reflected signals is 180°, is the combined signal stronger or weaker than the direct signal?

h. From what is the sky wave signal reflected?

i. Name three atmospheric factors on which the speed of a microwave signal depend.

j. Is the path of a microwave signal straight?

k. Under normal atmospheric conditions, does the path of a microwave signal curve upward or downward?

l. What is the propagation of radio waves called when atmospheric conditions are abnormal?

m. What type of nonstandard atmospheric conditions is likely to persist for any length of time?

n. What two abnormal atmospheric conditions are peculiar to microwaves?

o. Do absorption and scattering affect S-band operation particularly?

p. On a point-to-point communication system, with a transmission path lying over land, is the reflected signal important?

q. When a horizontally polarized wave strikes a horizontal surface, is the electric field parallel to the surface?

r. Does sea water reflect signal with the same strength as the incident signal? What is the phase shift of the reflected signal with reference to the incident signal?

s. Define angle of incidence as applied to a vertically polarized signal.

t. If the reflected signal of a point-to-point communication system over water arrives at the receiver precisely out of phase with the direct signal, is reliable operation of the system possible? How may the condition be remedied?

u. On a vertical polar diagram, two parts of the pattern are called *lobes* and *gaps*. In which of these can a radar pick up targets?

v. How does a radio line of sight differ from an optical line of sight?

w. The radius of the real earth is 4,000 miles. What is the radius of the modified earth? Why is the modified earth radius convenient to use on a modified earth-profile diagram?

x. On what kind of diagram is radar coverage usually shown?

y. If the radio line of sight between two points is obstructed, is communication possible?

z. If the frequency of a microwave communications system is doubled and the receiver antenna remains the same size, is the power density doubled or quadrupled?

CHAPTER 3

MICROWAVE TRANSMISSION SYSTEMS

Section I. COAXIAL LINE

45. Guided and Unguided Waves

When discussing the operation of radio and radar systems in the vhf region and at lower frequencies, it is customary to think of energy as being carried through space by electromagnetic waves and along transmission lines by voltages and currents. The voltage-current concept obviously is applicable in such conventional transmission systems as two-wire and coaxial lines, since both voltage and current can be measured readily. It is difficult, however, to apply the concept to a waveguide, which consists of a *single* hollow conductor on the outer surface of which neither voltage nor current can be detected. Although it is possible to understand the way a waveguide works in terms of voltages and currents, a clearer concept is gained by thinking of the energy as being carried by electromagnetic waves traveling inside the waveguide. The same approach, of course, can be applied to conventional transmission systems by considering the electromagnetic waves that accompany the voltages and currents on the line. In making use of the wave concept, it is desirable to distinguish between guided and unguided waves. Guided waves are those that carry energy along transmission lines, and unguided waves are those that carry energy through space. The path of a guided wave is fixed by the guiding structure; that of an unguided wave is determined by the propagation characteristics of the region through which it travels. The signals radiated from an antenna are unguided waves, even though the antenna may be highly directional because the antenna controls only the direction in which the waves are launched—not the direction of subsequent travel.

46. Types of Transmission Lines

The normal transmission systems used at microwave frequencies use either coaxial line or hollow waveguide. A coaxial line is smaller and lighter than a waveguide but it has higher losses, except at S-band frequencies near 3,000 mc per second where the losses in short sections of coaxial line are not much greater than those in hollow waveguides; normally, coaxial line is used for distances less than 15 feet, and waveguide for longer runs. At X- and K-band frequencies near 10,000 and 30,000 mc per second, the losses of coaxial line are far greater than those of waveguide, and, except for applications like coupling the local oscillator of a superheterodyne receiver to a waveguide transmission system, coaxial line seldom is used. The length of such a coaxial line normally does not exceed 1 or 2 inches. Two-wire transmission lines (and one-wire lines in which the ground serves as the second conductor) radiates strong signals, unless the spacing between conductors is much less than a wavelength of the signal being carried; such small spacing is difficult to achieve at microwave frequencies however, and lines of this sort are not used. In addition to hollow waveguide, which guides an electromagnetic wave along its interior, a type of single-conductor transmission line called surface waveguide, or G-line, sometimes is used at microwave frequencies. A G-line guides a wave that travels through the region surrounding the one-wire line, and, like waveguide, it requires no second conductor.

47. Use of Coaxial Line at Microwave Frequencies

Coaxial line has only two important applications at microwave frequencies; it is used as transmission line in the S-band over distances of approximately 15 feet or less, and in the X- and K-bands over distances of approximately 2 inches or less. Its merit for S-band operation lies in the fact that it is lighter and less bulky than waveguide and,

although its losses are greater than those of waveguide, the difference in loss is not significant for short runs. For longer runs, over 15 feet, waveguide generally is used. In the K- and X-bands, coaxial line is used to couple the output of low-powered oscillator to waveguide transmission lines, since it is fabricated and handled more easily than waveguide. However, in these bands, the losses of coaxial line are so much greater than those of hollow waveguide that it never is used for the main transmission lines of an installation.

a. S-Band Coaxial Line. To keep losses as low as possible, S-band coaxial line uses air as an insulator between the inner and outer conductors. The line usually is filled with dry air, and the pressure inside is held at approximately 5 pounds per square inch above atmospheric pressure, so that any leak results in the loss of dry air rather than an invasion of outside air, which may be moist. The use of conventional insulating material (other than air) is avoided by employing short sections of coaxial line to support the center conductor of the main line (fig. 27). If a transmission line is exactly one quarter-wavelength and is terminated in a short circuit, its input impedance

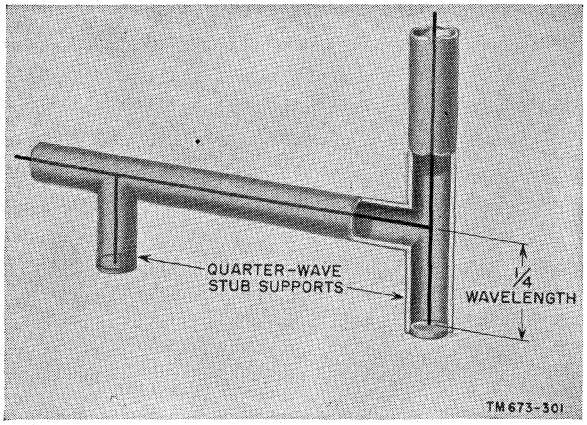


Figure 27. Coaxial line with quarter-wave stub supports.

is extremely high. The supports, called quarter-wave stub supports, are sections of this type. Their impedance at the operating frequency is so high that they do not affect the characteristics of the line, although at most other frequencies they look like low-impedance shunts across the line, and for direct current they look like a short circuit. Because of this type of construction, coaxial line is rigid, and joints must be provided wherever one section of line moves while an adjacent section is stationary.

b. X- and K-Band Coaxial Line. A typical use of coaxial line in the X- and K-bands is illustrated in figure 28, which shows a type 723A/B oscillator tube for X-band operation. The tube has a conventional octal base, and the coaxial line replaces one of the pins. The corresponding location at the socket is left clear so that the line can project through the socket and into the waveguide, as shown in figure 29. The coaxial line thus serves as a coupling betwen the tube and the transmission line. Here, the coaxial line consists of

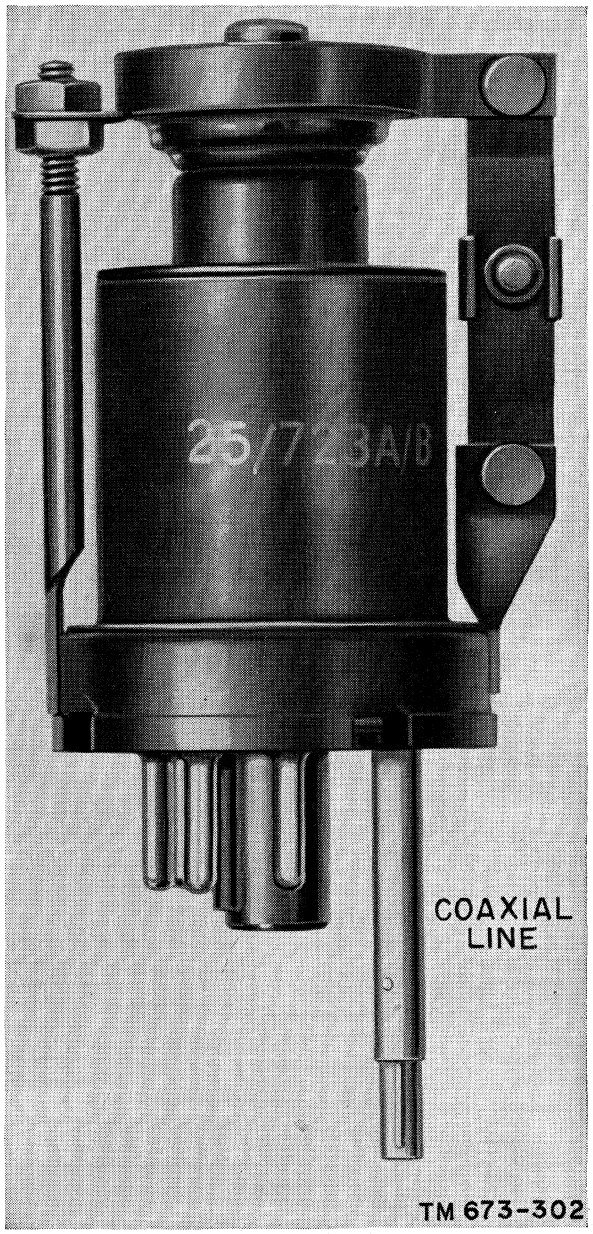


Figure 28. Coaxial line as part of X-band 723A/B oscillator tube.

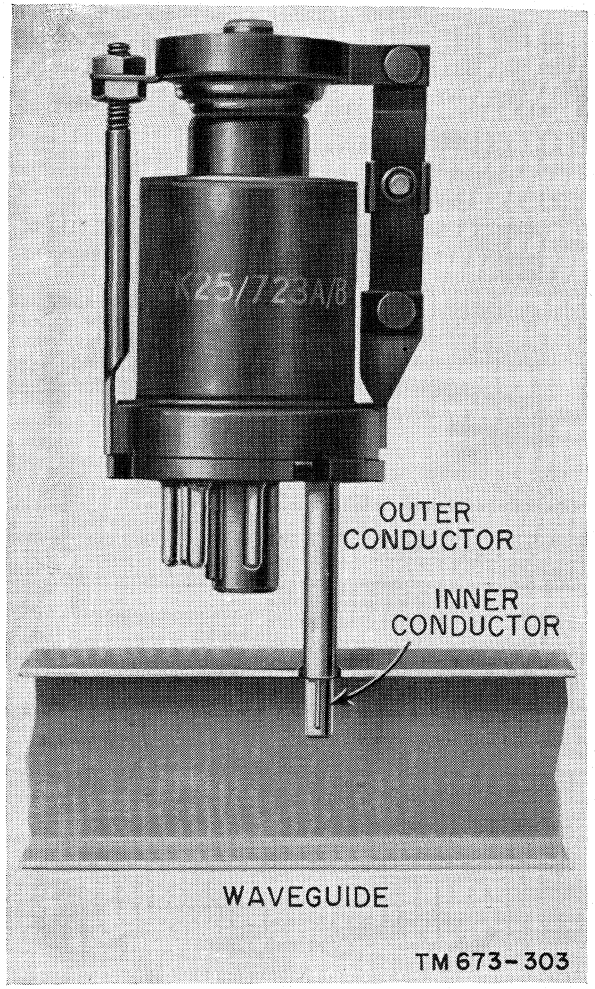


Figure 29. Coaxial-line coupling from 723A/B tube to waveguide.

silver inner and outer conductors with a solid insulator of polystyrene. The outer conductor is terminated at a point even with the upper surface of the waveguide, and the inner conductor is extended into the guide; the polystyrene insulator also is extended and serves to protect the rather fragile inner conductor. The extended inner conductor acts as an antenna and radiates energy into the space inside the waveguide. The losses in such a coaxial line are relatively high, but over such a short distance they are not important.

48. Electric and Magnetic Fields in Coaxial Line

Although the behavior of a conventional transmission line is explained adequately in terms of voltages and currents, the electric and magnetic fields associated with these should be examined in order to establish a background for the explanation of waveguide behavior in paragraphs 49–57. In *a–d* below, it is assumed that the line is a coaxial line having no losses.

a. Electric Field. At any point in the space between the coaxial-line conductors, an electric field is directed from the positive conductor to the negative conductor. At a given point, the strength of the field is proportional to the voltage difference between the conductors; for a given voltage, the strength of the field is inversely proportional to the distance between the axis of the line and the point in question. In A of figure 30, the arrows show the direction of the electric field when the outer conductor is positive and the inner conductor is negative. B shows the falling off of field strength as the distance from the center of the inner conductor increases; a stronger field is indicated by darker shading.

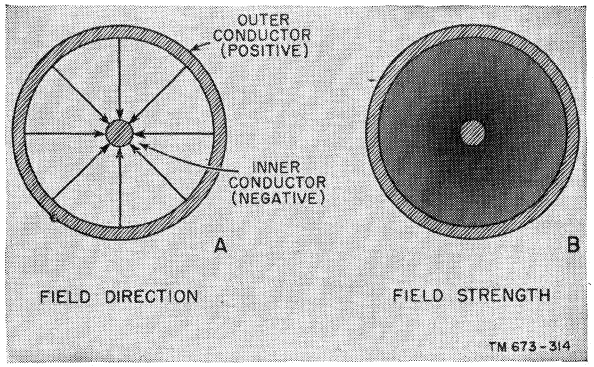


Figure 30. Coaxial-line electric field.

b. Magnetic Field. The direction of the magnetic field in a coaxial line is always tangent to a circle the center of which is on the axis of the line, as shown in A of figure 31. At a given point, the strength of a magnetic field is proportional to the

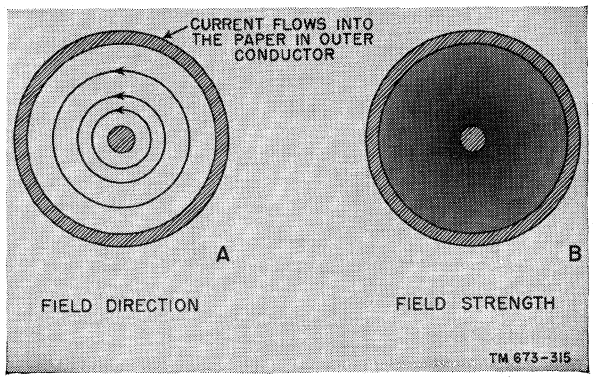


Figure 31. Coaxial-line magnetic field.

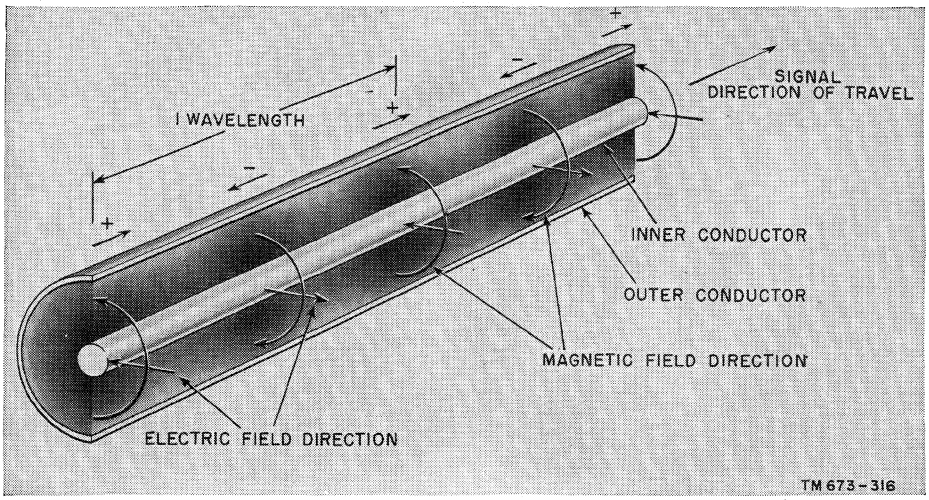


Figure 32. Coaxial-line combined electric and magnetic fields.

current flow; for a given current, the strength of the magnetic field varies inversely with the distance from the axis of the line, as shown in B (just as the strength of electric field varies). Reversing the direction of the current causes a reversal in the direction of the magnetic field.

c. Combined Fields. Figure 32 shows the combined electric and magnetic fields in a coaxial line that is carrying a signal in only one direction. The line is shown cut in half along its axis, and the plane along which it is cut has been shaded to show the strength of the fields. The signs and arrows above the line show the sign of the voltage on the outer conductor and the direction of current flow in the outer conductor. The straight arrows to the right of the shaded cut section show the direction of the electric field, either toward or away from the center conductor. The curved arrows to the right show the direction of the magnetic field, either clockwise or counterclockwise around the inner conductor. The line was assumed to be lossless, so that the characteristic impedance Z_o is a pure resistance. Voltage and current are then in phase, and peak voltage and peak current at a given point occur at the same time. Similarly, both voltages and current vanish simultaneously at a given point. Since E equals IR at all points on the line (where R is the resistance that is equal to Z_o), the voltage is always proportional to the current and the electric field is always proportional to the magnetic field. The same shading then serves to show the strength of both fields.

d. Electromagnetic Wave. A of figure 33, shows the strength and direction of the electric and magnetic fields for points lying on a line parallel to the path of an electromagnetic wave traveling through space. The direction of the electric field or the magnetic field at any point is shown by the direction of the arrow the tail of which lies at the point in question. (One arrow shows the electric

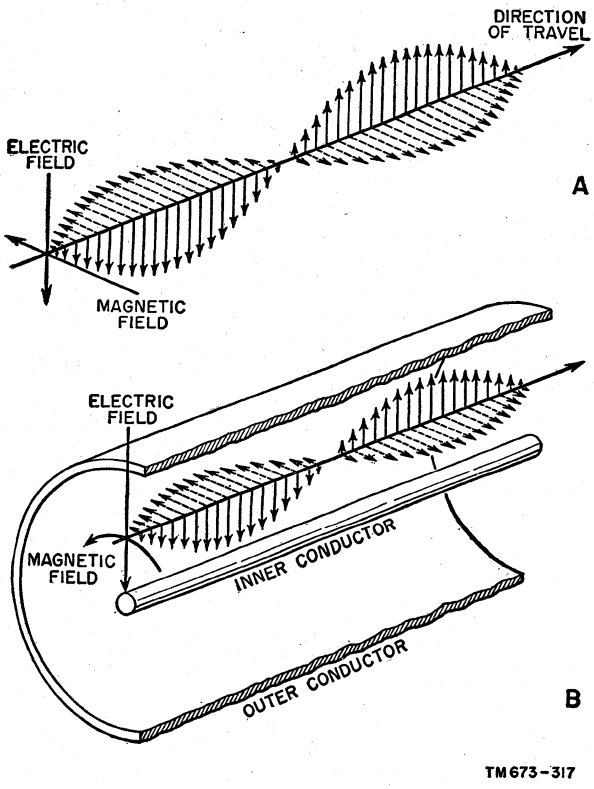


Figure 33. Electromagnetic waves in space and in coaxial line.

field; the other shows the magnetic field.) The strength of the field is shown by the length of the arrow. In the same way, B shows the electric and magnetic fields for points on a line parallel to the axis of a coaxial transmission line. From the similarity in the two field arrangements, it can be seen that the fields in the coaxial line form an electromagnetic wave which travels through the space between the inner and outer conductors in exactly the same way that an electromagnetic wave travels through unobstructed space. It is, therefore, easy to think of energy as being carried down the coaxial line by an electromagnetic wave (although, of course, it is even easier to think of it as being carried by voltage and current). In the following section, the electromagnetic wave concept is used to explain how energy is carried through the space inside a hollow waveguide.

Section II. WAVEGUIDES

49. Types of Waveguide

Although any transmission system serves as a guiding structure for guided waves, the term *waveguide* is reserved for the type that consists of a single hollow conductor. Although a waveguide may have almost any shape, the two types in common use are those having rectangular and circular cross sections (fig. 34). The characteristics of the rectangular waveguide shown in A are determined by the height and width. These dimensions are always the inside measurements, the smaller of the two customarily being taken as the height, even though it may lie in a horizontal plane. The characteristics of the circular waveguide shown in B are determined by the diameter, which is always the inside measurement.

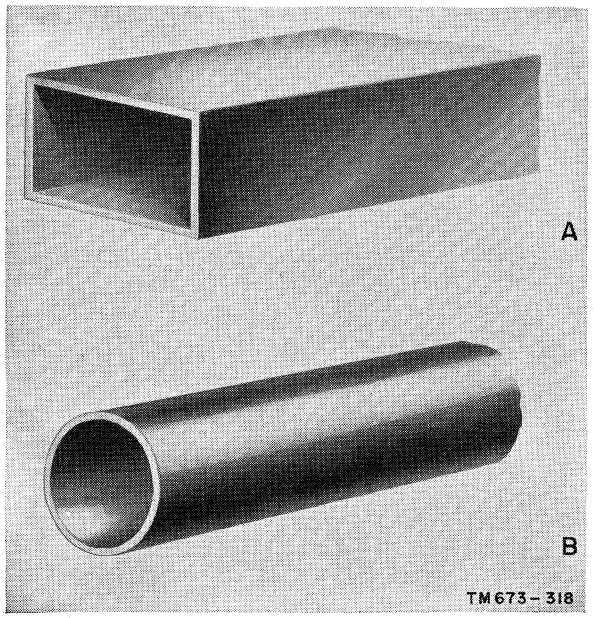


Figure 34. Rectangular and circular waveguides.

50. Transmission in Rectangular Waveguide

Although a rectangular waveguide can carry energy in many ways, only one way, the simplest, is usual in practice. In this case, vertically polarized waves travel down the waveguide along paths that are parallel to the top and bottom of the guide.

a. Terminology of Rectangular Waveguide. As stated in the preceding paragraph, the electrical characteristics of a rectangular waveguide are determined by the inside dimensions (fig. 35).

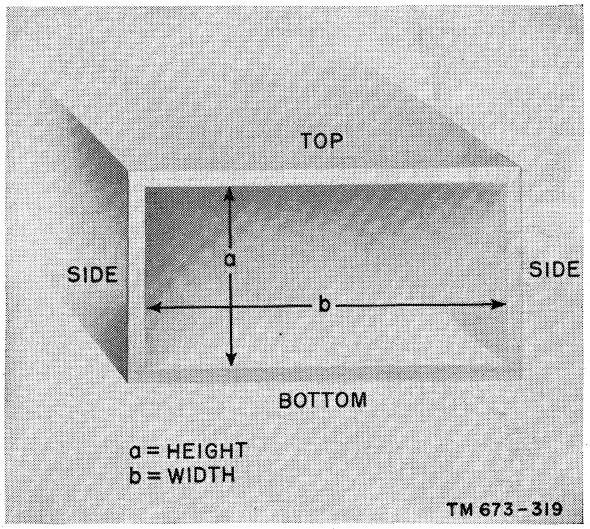


Figure 35. Rectangular waveguide dimensions.

Whatever the orientation of the waveguide, it is customary to call the shorter dimension the height, and the longer one the width. Following this thought, the longer walls of the waveguide are called the top and bottom, and the shorter walls are called the sides. *Vertical polarization as applied to an electromagnetic wave in a rectangular waveguide, therefore, means that the electric field is*

parallel to the shorter dimension of the guide. The small letters *a* and *b* customarily are used to denote the height and width, respectively, when referring to rectangular waveguide dimensions.

b. Paths of Waves. The paths followed by the electromagnetic waves are shown in figure 36. A is a side view, with the wave paths parallel to the top and bottom walls of the waveguide. B is a top view of the paths. All of the paths make the same angle, B, with the sides of the waveguide.

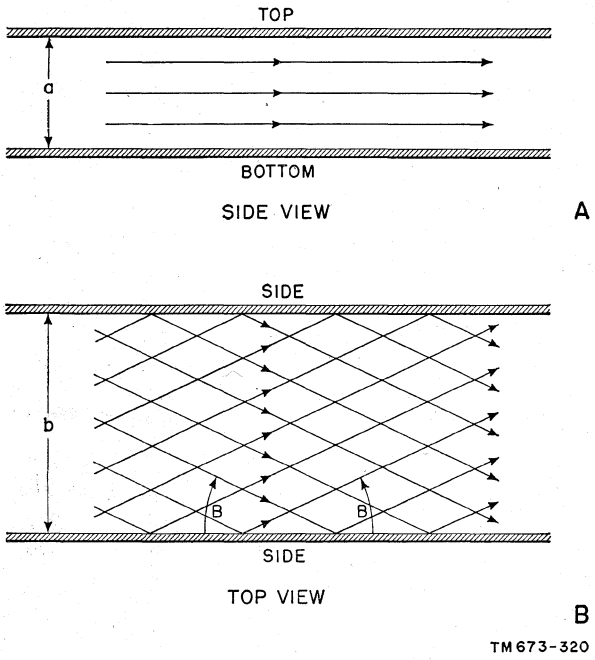


Figure 36. Paths of waves in rectangular waveguide.

c. Determination of Angle B. Angle B cannot be chosen arbitrarily. Instead, it depends on *b*, the width of the waveguide, and on λ, the wavelength of the signals. The relationship is given by the equation,

$$\sin B = \lambda/2b.$$

This equation results from the fact that propagation is impossible unless all signals traveling in a given direction reinforce each other. The equation is derived easily from consideration of the diagram shown in figure 37. Consider a signal that passes through C moving to the left. It is reflected from the left-hand wall at D, from the right-hand wall at E, and finally moves in direction EF, which is parallel to the original direction, CD. A wave front is always perpendicular to the path of a wave, and since DH is perpendicular to CD, DH is a wave front for signals moving to the left. As explained in paragraph 17*b*, the phase of a signal is the same at all points on a wave front. A signal moving to the left through H is, therefore, in phase with the incident signal at D. If energy is to be propagated down the waveguide, the reflected signal at E must have the same phase as the signal moving to the left through I, since both are on a wave front perpendicular to EF. Otherwise, the signals will cancel each other after traveling a short distance through the waveguide. Since the electric field of the signals is parallel to the reflecting wall, there is a phase reversal at D and another one at E. The signal behavior on reflection is the same as that of a horizontally polarized signal reflected from the surface of the sea, as described in chapter 2. However, it does not matter whether this is the case, since two phase reversals have a net effect of zero. *To have the desired phase relations, angle B must have a value*

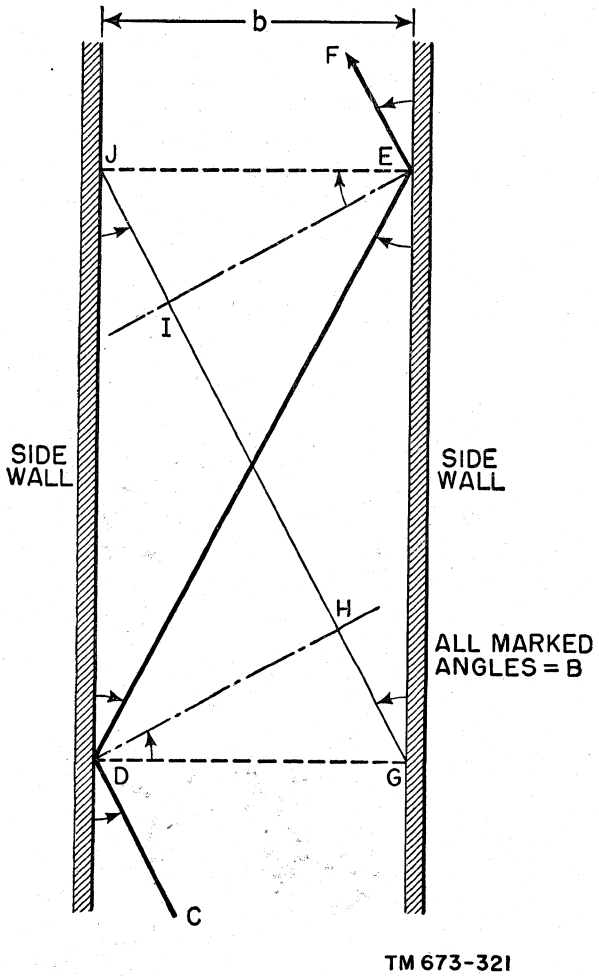


Figure 37. Geometry of equation for determination of angle B.

which makes path DE an integral (whole) number of wavelengths longer than path HI. GJ is equal to DE, since G lies opposite D, and J lies opposite E. The requirement for propagation of energy is, then, that HI and GJ differ by an integral number of wavelengths. The difference between GJ and HI is the sum of GH and IJ, and each of these distances is equal to $b \sin B$. The condition for propagation, then, is that

$$2b \sin B = n\lambda$$

where n is an integer (1, 2, 3, and so on). In the simplest case, which is the one under discussion, n is equal to 1 and the equation given above immediately reduces to

$$\sin B = \lambda/2b,$$

when 1 is substituted for n. It must be emphasized that the waveguide will carry energy only for certain values of angle B, because the signals cancel for other values of that angle. This has nothing to do with attenuation or losses; it is similar to the directive effect of an antenna that does not radiate in directions for which signals from various portions of the antenna cancel each other.

51. Field Patterns in Rectangular Waveguide

The patterns of the electric and the magnetic fields in the rectangular waveguide may be found by adding the fields belonging to the two sets of electromagnetic waves—those moving to the left and those moving to the right. This may be done separately for the electric and the magnetic fields.

a. Spacing of Wave Fronts. To construct the field patterns, it is necessary to know the spacing in the waveguide between two wave fronts having a phase difference of 1 half-cycle. This is developed in the diagrams shown in figure 38. Here CD is a line drawn across the waveguide and CE is a wave front of the signal that is moving to the left. The discussion in paragraph 50*c*, has shown that DF (which corresponds to GH in fig. 37) must be 1 half-wavelength in the simple case (for which n is 1). Wave front DG in figure 38 is therefore 1 half-cycle different in phase from wave front CE. A shows a fairly wide waveguide, and B shows a narrow waveguide.

b. Pattern of Electric Fields. Figure 39 illustrates the way in which the electric fields combine. A shows the electric fields of the waves

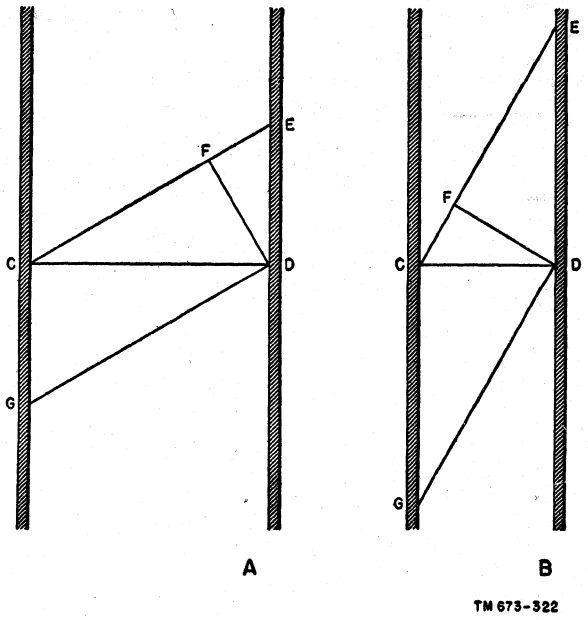


Figure 38. Spacing of wave fronts.

that are moving to the left, and B shows those of the waves moving to the right. In each illustration, the direction of the electric field is indicated by the notes beside the figure, and the strength of the field is shown by the darkness of the shading. C shows the combined fields that are obtained simply by adding the fields of the two sets of waves.

c. Pattern of Magnetic Fields. Figure 40 illustrates the way in which the magnetic fields combine. When adding these fields, it is necessary to consider the relative directions of the individual fields, in the same way that, when adding two a-c voltages, it is necessary to consider the relative phases. The direction of the magnetic field is shown, in A and B, by small arrows next to the waveguide, and in C, by arrows superimposed on the shading. The strength of the fields is indicated by the darkness of the shading.

d. Combined Pattern of Electric and Magnetic Fields. The combined pattern of the electric and magnetic fields is shown in figure 41. Here, the strength of the electric field is indicated by the darkness of the shading, and the direction is given as a note beside the waveguide. The character of the magnetic field is indicated by the closed loops. The arrowheads on these loops show the direction of the magnetic field, and the spacing of the loops indicates the strength of the magnetic field; closely spaced loops indicate a strong field. The magnetic field pattern is shown more pre-

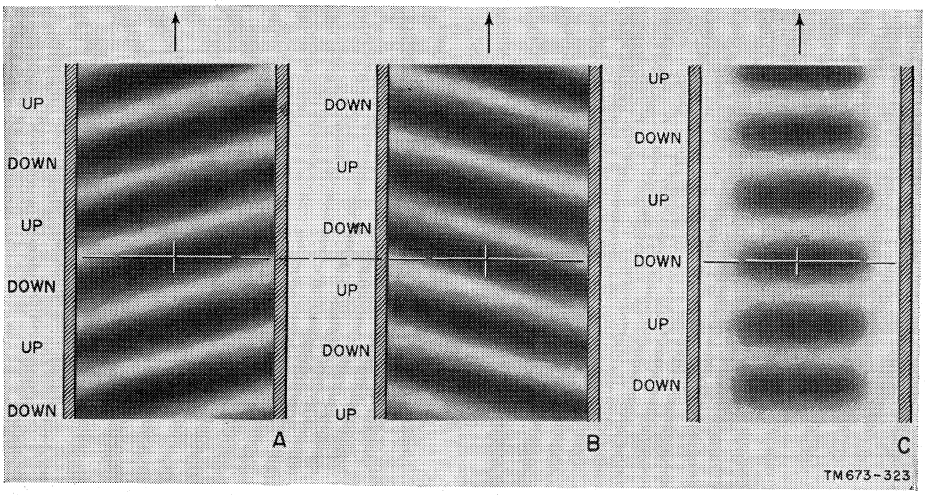


Figure 39. Pattern of electric fields.

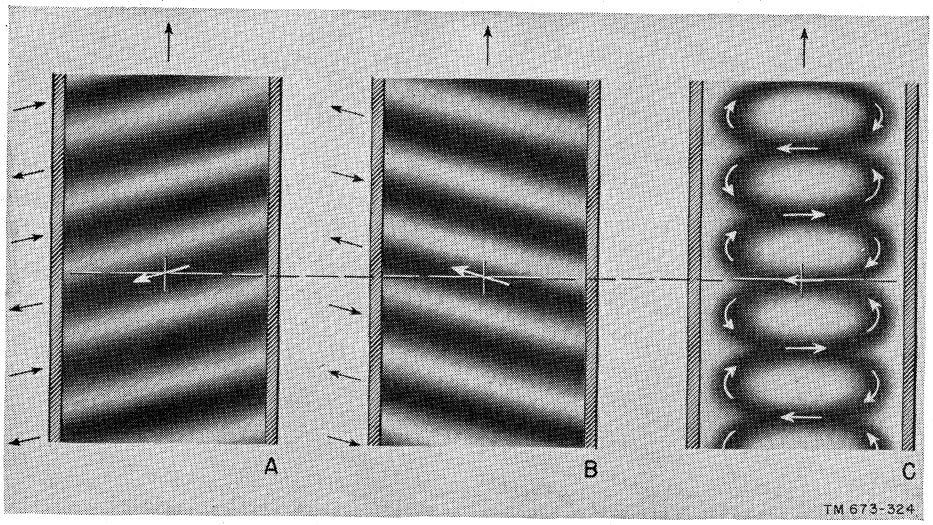


Figure 40. Pattern of magnetic fields.

cisely in C of figure 40. At both side walls, the electric field vanishes and the magnetic field is longitudinal (parallel to the axis of the waveguide). Midway between the side walls, the field is the same as in unobstructed space, both electric and magnetic fields being transverse (perpendicular to the waveguide axis). At intermediate points, fields of both sorts are present with reduced strength, the longitudinal and transverse magnetic fields adding to produce a combined magnetic field that is neither transverse nor longitudinal.

e. Details of Combined Pattern. At any point inside the waveguide, sinusoidally oscillating fields will be observed as the field pattern travels down the guide past the point of observation. The character of the observed fields depends only on the distance between the point of observation and the side wall of the waveguide. At a point midway between the side walls of the waveguide, the electric field, A, and the transverse magnetic field, B, vary, as shown in figure 42. The two fields are in phase with each other, the peak intensity of the electric field occurring at the same instant as the peak intensity of the transverse magnetic field. No longitudinal magnetic field is observed at a point midway between the side walls. At a point somewhere between the center of the guide and one of the side walls, the fields vary, as in A and B of figure 43. This figure shows the fields for a point between the center of the guide and the right-hand side wall. For a

point between the center of the guide and the left-hand wall, the observed fields are the same as those in figure 43 except that the phase of the

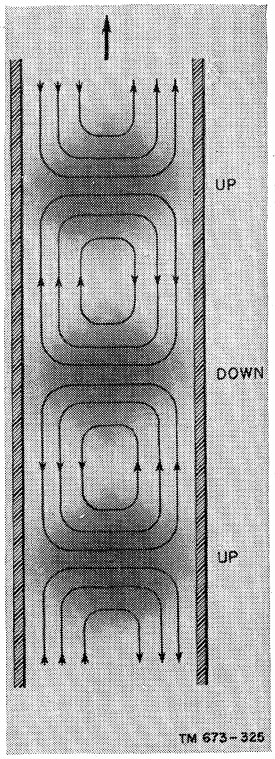


Figure 41. Combined pattern of electric and magnetic fields.

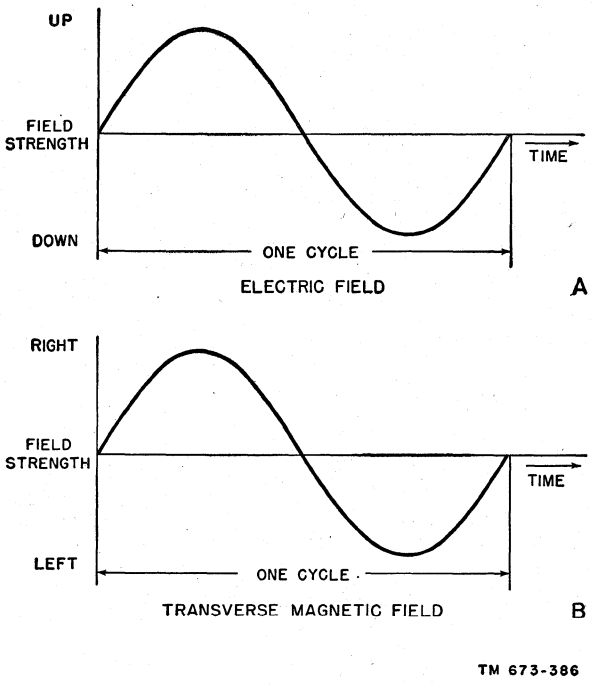


Figure 42. Field observed midway between side walls.

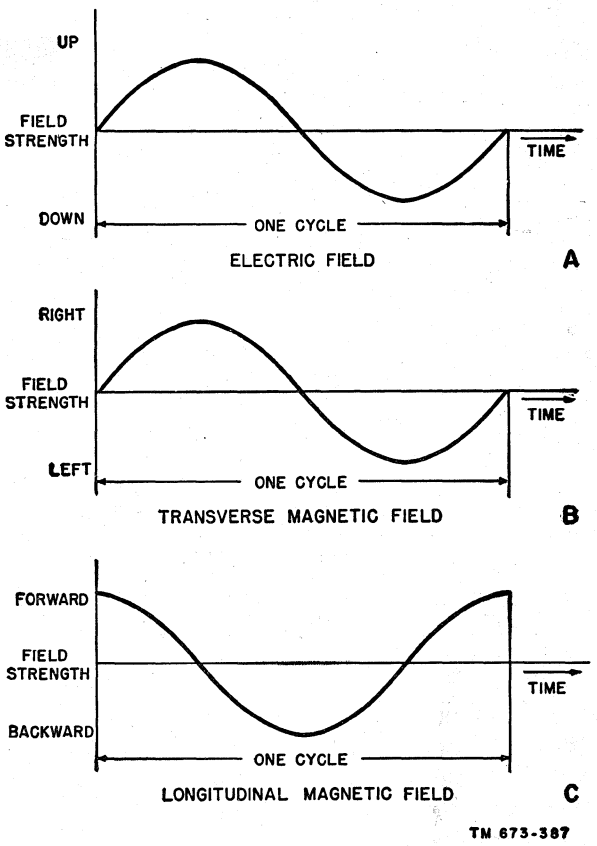


Figure 43. Fields observed at point between center of guide and right-hand wall.

longitudinal magnetic field, in C, is reversed. At both side walls, the longitudinal magnetic field is present but both the transverse magnetic field and the electric field vanish. If the amplitude of the oscillating field is plotted against distance from the side wall of the waveguide, the resulting curve is half of a sine wave (fig. 44). For this reason, the pattern sometimes is called a *half-sine* pattern. The electric and the transverse magnetic fields vanish at both walls, as in A. The longitudinal magnetic field vanishes at the center of the guide, as in B, and it has opposite phases in the right and left halves of the guide, as shown by the negative value of the amplitude in the figure.

52. Cut-Off Wavelength

As explained in paragraph 50*c*, angle B is determined by the relationship,

$$\sin B = \lambda/2b$$

for the simplest wave pattern that will carry energy down the waveguide. This equation shows that B

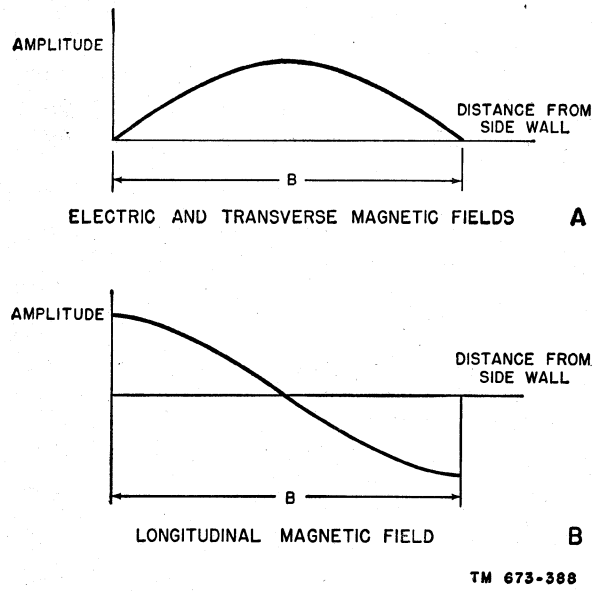


Figure 44. Field amplitudes.

must equal 90° when the wavelength is equal to $2b$. Here, the two sets of waves merely travel back and forth across the waveguide and no energy is carried in the desired direction (along the guide). If the wavelength is greater than $2b$, there is no value of B that satisfies the relation, and a wave pattern of the sort described in paragraph 51 cannot exist. The wavelength at which B equals 90° is called the λ_c (*cut-off wavelength*), and the corresponding frequency is called the f_c (*cut-off frequency*). For the wave pattern described above, λ_c is equal to $2b$. To avoid confusion, the wavelength of the signal is referred to as the λ_a (*wavelength in air*). (A signal with a frequency of 3,000 mc per second, for example, has λ_a equal to 10 centimeters.) Using this terminology, a waveguide cannot transmit energy by means of the simple pattern described in paragraph 51 *unless* λ_a is shorter than λ_c. Waveguide transmission sometimes is explained in terms of a two-wire transmission line with an infinite number of quarter-wave shorted stubs, each one of which has an infinite impedance and therefore looks like an open circuit. Figure 45 shows a two-wire line with a large number of stubs, and the similarity to a waveguide is immediately apparent. This explanation does not cover the field pattern nor the group and phase velocities, which are discussed in paragraph 54. It implies, further, that transmission is possible *only* if the width of the guide is 1 half-wavelength, which is incorrect since transmission is possible only when the width of the guide *exceeds* 1 half wavelength. It may be useful to remember, however, that a waveguide at cut-off frequency (where it will not carry energy) looks like a two-wire line with an infinite number of quarter-wave shorted stubs.

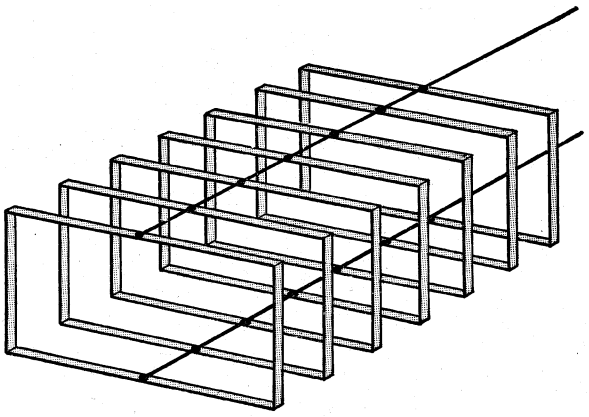


Figure 45. Two-wire transmission line with quarter-wave shorted stubs

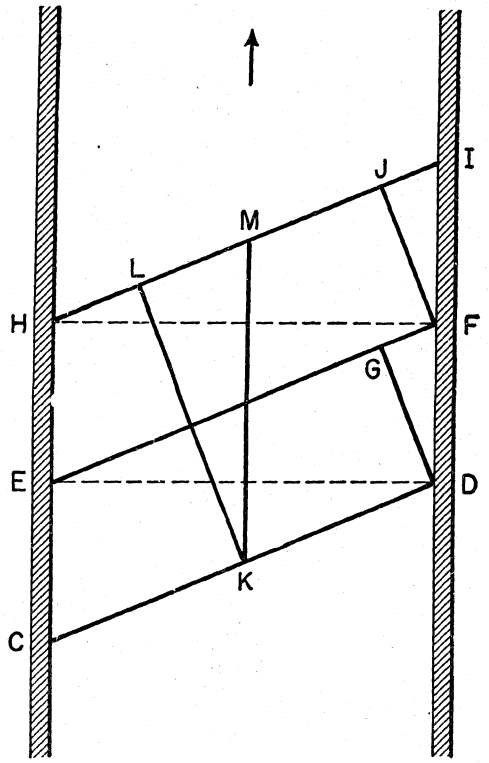
53. Guide Wavelength

In addition to λ_a and λ_c, it is customary to refer to the λ_g (*guide wavelength*). This is the wavelength of the pattern in the waveguide, and it is equal to $\lambda_a/\cos B$. The relationship is derived easily from the diagram in figure 46. Here, CD is a wave front of the signal that moves to the left. E is on the left-hand wall directly opposite D, and EF is another wave front of the same signal. Distance DG is equal to $b \sin B$ and, from the discussion in paragraph 50*c*, it is clear that this must be $\lambda_a/2$. The two wave fronts, CD and EF, are, therefore, separated by one-half λ_a. H is on the left-hand wall directly opposite F, and HI is another wave front. By an argument similar to the one given above, the spacing between EF and HI is one-half λ_a. Then KL, the spacing between CD and HI, is exactly equal to λ_a. The wavelength of the pattern, however, is equal to KM, since this is the distance along the guide axis between two points where the phase of the signal differs by 1 cycle. The angle between KL and KM is B, whence

$$KL=\lambda_a=KM \cos B=\lambda_g \cos B,$$

and

$$\lambda g=\frac{\lambda_a}{\cos B}$$



TM 673-327

Figure 46. Geometry of equation for derivation of guide wavelength.

54. Group and Phase Velocities

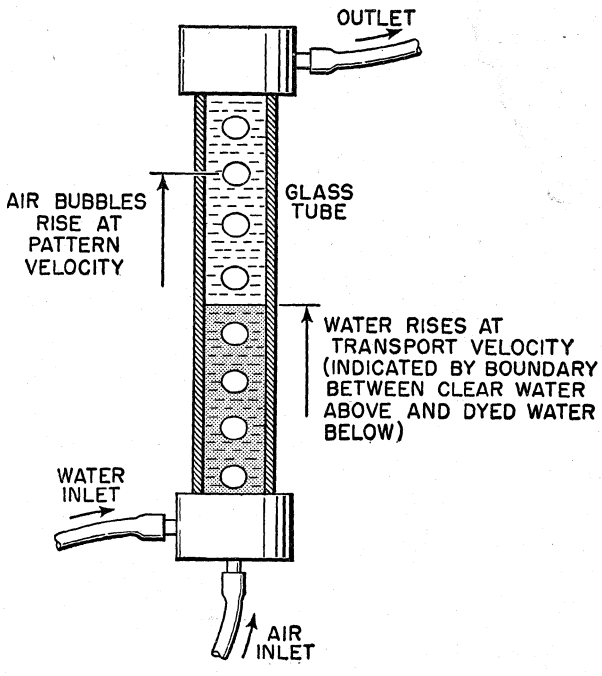
Just as two wavelengths, λ_a and λ_g, must be considered in connection with the transmission of energy along a waveguide, three separate velocities are of concern. The first of these is the velocity of electromagnetic waves in space, usually written c, which is equal to 330 yards or 300 meters per microsecond. The other two are the v_g (*group velocity*), the velocity with which energy is transmitted along the waveguide, and the v_p (*phase velocity*), the velocity with which the field pattern moves along the waveguide.

a. Distinction Between Velocities. Imagine a transmission system, similar to the one shown in figure 47, that consists of a vertical glass tube through which water is pumped. Simultaneously, air is introduced at the bottom and rises in the form of small, regularly spaced bubbles. Suppose, now, that clear water is pumped through the tube. The rising bubbles will form a pattern which moves upward at a speed that may be called the *pattern velocity*. If dyed water then is introduced, the boundary between the clear water and the colored water will rise at a speed that may be called the *transport velocity*—the velocity at which water moves through the tube. The case is similar to that of a waveguide. If a steady signal is sent through the guide, the field pattern will have a wavelength, λ_g, analagous to the spacing between the bubbles. The field pattern will move with a phase velocity, just as the bubble pattern moves at the *pattern velocity*. If, then, the signal level is raised suddenly, the boundary between the low-level signal and the high-level signal will move along the waveguide with a group velocity, just as the boundary between the clear water and the dye solution moves at the *transport velocity*.

b. Variation of Velocities With Wavelength. As explained in paragraph 52, the cut-off wavelength, λ_c, is equal to $2b$, and angle B is determined by the relation,

$$\sin B = \lambda_a/2b = \lambda_a/\lambda_c$$

When the operating frequency is extremely high, so that the wavelength of the signal (in air) is very much shorter than the cut-off wavelength of a waveguide, sin B and B are both very small and cos B is nearly 1. In this case, the phase velocity is only slightly greater than c and the group veloc-


OUTLET
AIR BUBBLES RISE AT PATTERN VELOCITY
GLASS TUBE
WATER RISES AT TRANSPORT VELOCITY (INDICATED BY BOUNDARY BETWEEN CLEAR WATER ABOVE AND DYED WATER BELOW)
WATER INLET
AIR INLET

TM 673-328

Figure 47. Pattern and transport velocities.

ity is only slightly smaller. At the other extreme, when the wavelength in air is only slightly shorter than the cut-off wavelength, sin B is very nearly 1 and B is nearly 90°. Cos B is then very small, so that v_g becomes small and v_p becomes extremely large. Similar behavior can be observed at any straight beach. Figure 48 shows a series of wave crests moving in a direction that makes an angle X with the edge of the beach. As a point on one of the wave crests moves from P to Q, the intersection of that crest with the beach moves from R to Q. By simple geometry, distance PQ is equal to distance RQ times cos X. Then, if the waves move forward with a velocity v_w, and the point at which the wave crest breaks on the beach moves along the edge of the beach with a velocity v_b,

$$PQ=v_wt, \quad RQ=v_bt,$$

(where t is the time required for the wave crest to move from P to Q), and

$$v_wt=v_bt \cos X,$$

so that

$$v_b=v_w/\cos X$$

The point of intersection of a wave crest with the edge of the beach, therefore, runs along the beach $1/\cos X$ times as fast as the wave pattern moves.

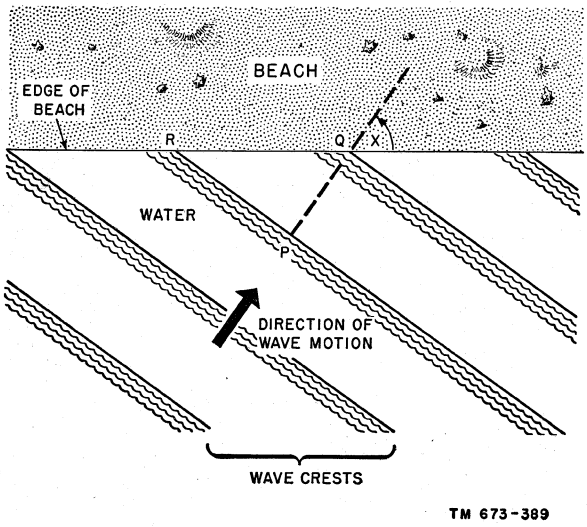


Figure 48. Water waves breaking on beach.

55. Other Modes of Transmission

A number of individual electromagnetic waves can combine in an infinite number of ways to carry energy down the interior of a rectangular waveguide. The various ways are called *modes*, and each mode has a distinctive field pattern different from that of every other mode. Each mode also has its own cut-off frequency, cut-off wavelength, guide wavelength, group velocity, and phase velocity, not necessarily different from those of other modes. In a rectangular waveguide, it is customary to choose dimensions that make the cut-off frequency higher than the operating frequency for all modes except the simple one used as an example in paragraphs 51–54. For all modes except this one, the operating frequency is below the cut-off frequency (and the operating wavelength is longer than the cut-off wavelength) so that transmission is impossible. The waveguide, therefore, operates only in the simple mode. In spite of this practice, the other modes should be considered briefly.

a. Mode Terminology. In some modes, the electric field is always transverse (where it does not vanish), whereas the magnetic field is sometimes transverse and sometimes longitudinal. This is true, for example, with the simple mode described above. Such modes are called *TE* (*transverse electric*) modes. In other modes, the magnetic field is always transverse, whereas the electric field is sometimes transverse and sometimes longitudinal. These are called *TM* (*transverse magnetic*) modes. In still other modes, both the electric field and the magnetic field are longitudinal at various points. These modes are not named, since a mode of this sort always can be considered as a combination of a given *TE* mode with a given *TM* mode. A mode is defined further by two subscripts, m and n, as a $TE_{m,n}$ or a $TM_{m,n}$ mode. The first subscript is the number of times that a pattern repeats itself along the a dimension of the waveguide (from the top of the guide to the bottom) and the second subscript is the number of times that the pattern repeats itself along the b dimension (from one side of the waveguide to the other). In the simple case described earlier, the field pattern is uniform from top to bottom, so that there is *no* repetition and the first subscript is therefore 0. There is variation, however, from one side to the other; but the pattern occurs only once, and therefore the second subscript is 1. The simple mode is, accordingly, called the $TE_{0,1}$ mode (since the electric field is entirely transverse). The following discussion shows that the $TE_{0,1}$ mode has a longer cut-off wavelength than any other mode for a given rectangular waveguide. Because of this, it is possible to choose the waveguide dimensions so that λ_c is longer than λ_a for this mode, and shorter than λ_a

for every other mode. The waveguide is forced thereby to operate in the $TE_{0,1}$ mode, as explained above.

b. Two Nonexistent Modes. It is impossible to send a single vertically polarized wave down a waveguide along a path that is parallel to the axis of the guide. The field pattern of such a wave will have an electric field parallel to each side wall *at the surface* of the wall. This will cause currents to flow in the side walls, and the electromagnetic waves radiated by these currents will cancel the original signal in a very short distance. The same thing is, of course, true of a horizontally polarized wave that travels *directly* down the guide without reflection. Here, there will be electric fields parallel to the top and bottom of the guide *at* the top and bottom, and the same result will be obtained. Finally, any electromagnetic wave may be considered as a combination of vertically and horizontally polarized waves. It follows, therefore, that it is impossible to send energy through a rectangular waveguide by means of any signal that travels directly down the guide without reflection. Now, the $TE_{0,0}$ and $TM_{0,0}$ modes have field patterns that are uniform along both the a and the b dimensions. Paragraph 50 explains that reflection from the side walls causes the field pattern to vary along the b dimension, whereas the absence of reflection from the top and bottom of the waveguide makes the pattern uniform along the a dimension. Thus, a field pattern uniform along the a and b dimensions can be obtained only if there is no reflection from any of the walls. It is, however, impossible to transmit energy through a rectangular waveguide by means of waves that are not reflected, and for this reason *neither the $TE_{0,0}$ nor the $TM_{0,0}$ mode can exist in a rectangular guide.*

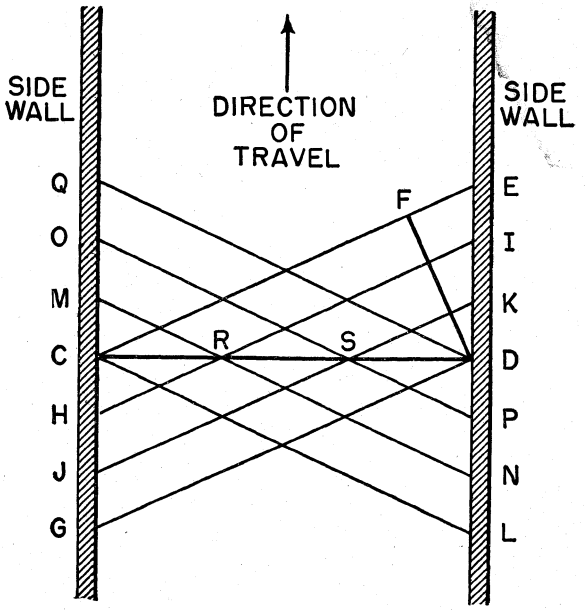
c. $TE_{0,n}$ Modes. It has been stated that energy can be carried along a waveguide by vertically polarized waves traveling parallel to the top and bottom of the guide if their paths make the proper angle with the side walls; also, that the proper angle is given by the equation,

$$2b \sin (B) = n\lambda_a,$$

or by

$$\sin (B) = n\lambda_a/2b,$$

where B is the angle between the paths of the individual waves and the side walls. It has been shown that the $TE_{0,1}$ mode is produced when n is equal to 1, although the mode was not so labeled at the time. For any value of n (except zero), this arrangement of vertically polarized waves reflected from the side walls produces the $TE_{0,n}$ mode. This may be verified by examination of the diagram shown in figure 49, which treats the case of n equal to 3. Here C and D are two points opposite each other on the side walls of a rectangular waveguide. The distance DF is n half-wavelengths (3 in this case) and wave front GD is, therefore, 3 half-wavelengths behind the wave front CE. It is, then, possible to construct additional wave fronts HI and JK, so that each is 1 half-wavelength from its neighbors. In the same way, the wave fronts QD, OP, MN, and CL may be constructed, each 1 half-wavelength from its neighbors. Now, as one moves from C to R, the phase of the waves moving to the left changes ½ cycle in one direction, and the phase of the waves moving to the right changes an equal amount in the other direction. This leads to a pattern similar to that obtained by moving *all the way* across the guide in the $TE_{0,1}$ mode (figs. 38 through 41). Conditions at R are the same as those at C, except that the phases of both sets of waves are reversed. The pattern between R and S is, therefore, the same as the pattern between C and R, but the directions of the fields are reversed. The same thing happens between S and D, with the directions of the fields the same as those



TM 673-331

Figure 49. Geometry of $TE_{0,3}$ mode.

between C and R. Consequently, the pattern for a $TE_{0,n}$ mode is nothing more than a repetition of the $TE_{0,1}$ pattern, with the field directions reversed at each repetition. Figure 50 illustrates the electric field in a plane perpendicular to the axis of a rectangular waveguide; A shows the $TE_{0,1}$ mode, B shows the $TE_{0,2}$ mode, and C shows the $TE_{0,3}$ mode. In B, the direction of the electric field in the right half of the waveguide is opposite to that in the left half. In one half, the field is directed upward; in the other half it is directed downward. Similarly, in C, the direction in the center of the guide is opposite to the direction in either of the side portions of the guide. In every case, however, the field is either upward or downward. Figure 51 shows the combined fields, as in figure 41, for the $TE_{0,3}$ mode. Other mode characteristics are discussed in appendix III.

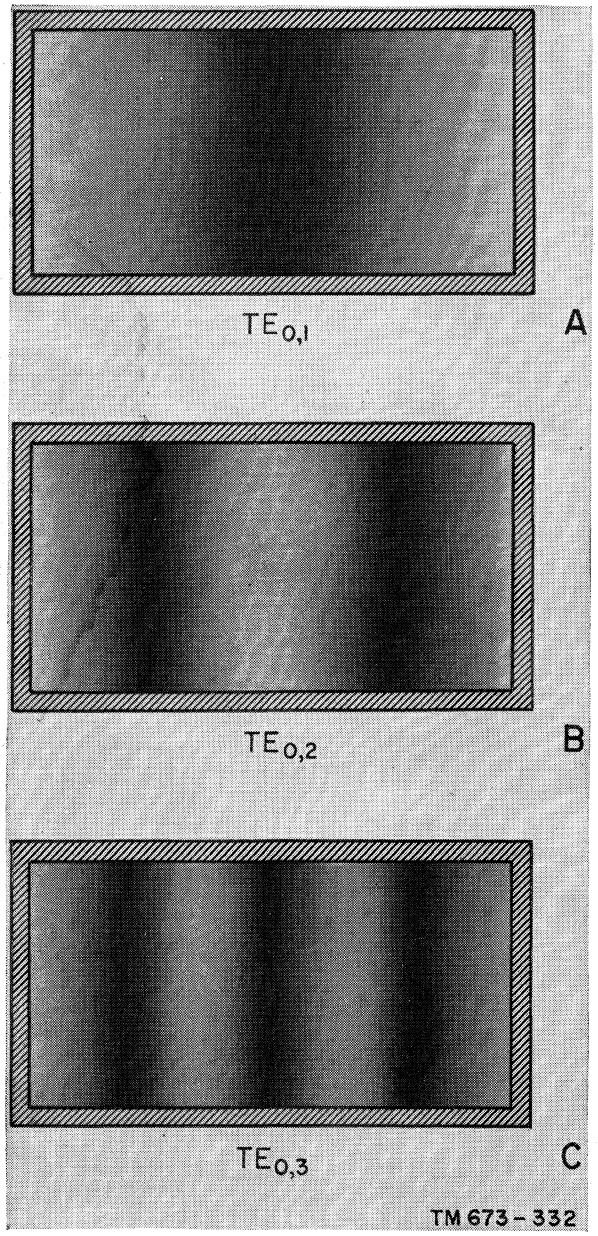


Figure 50. Electric field patterns for $TE_{0,n}$ modes.

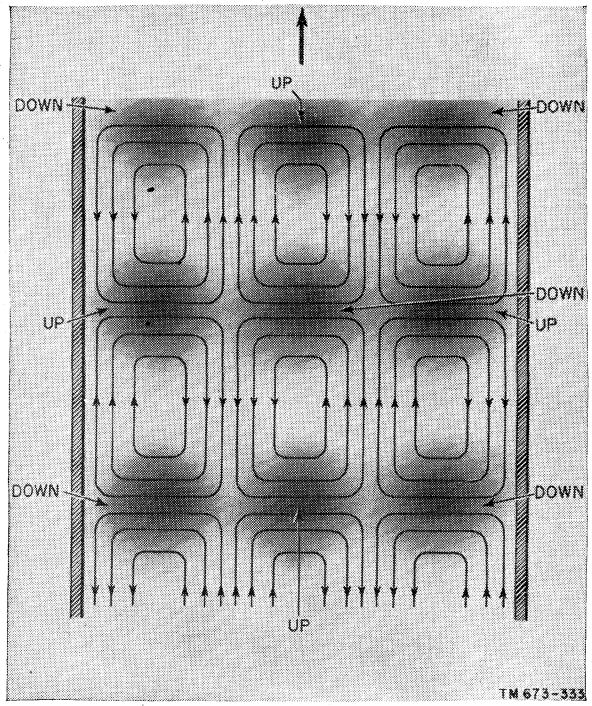


Figure 51. Combined fields for $TE_{0,3}$ mode.

d. Other Nonexistent Modes. At first glance, it would appear possible to develop the $TM_{0,n}$ mode by sending horizontally polarized signals along the paths prescribed for the signals that produce the $TE_{0,n}$ mode. This will produce a pattern of the desired type, but there will be longitudinal electric fields at the walls and the pattern will be canceled by the waves radiated by induced wall currents. The same objection applies to the possibility of developing a $TM_{m,0}$ mode by sending vertically polarized waves along the paths that produce a $TE_{m,0}$ mode. For this reason, *neither the $TM_{0,n}$ modes nor the $TM_{m,0}$ modes can exist in a rectangular waveguide.*

e. Reflection From Four Walls. All of the modes previously considered have patterns that are built up from two sets of waves, and, before considering the $TE_{m,n}$ and $TM_{m,n}$ modes, it is necessary to consider what happens when the signals are reflected from all four walls of the waveguide. The pattern then will be built up from four sets

of waves, those moving downward to the right and left, and those moving upward to the right and left. To begin, consider the two sets of waves that move downward to the right and left. Their paths meet the side walls of the waveguide at some angle B which satisfies the equation

$$\sin B = n\,\lambda_a/2b,$$

and the two sets of waves combine to form a field pattern like that of the $TE_{0,n}$ mode, *except* that it is tipped downward and travels downward. A similar pattern, tipped in the opposite direction and traveling upward, is formed by the two sets of waves that move upward to the right and left. These two patterns follow paths that meet the top and bottom of the waveguide at an angle A. In figure 52, A shows the paths from above and B shows them from the side. Since each pattern is composed of two sets of electromagnetic waves which are reflected by the top and bottom of the waveguide, the pattern as a whole is reflected, and it behaves, generally, like one of the sets of waves in a $TE_{m,0}$ mode.

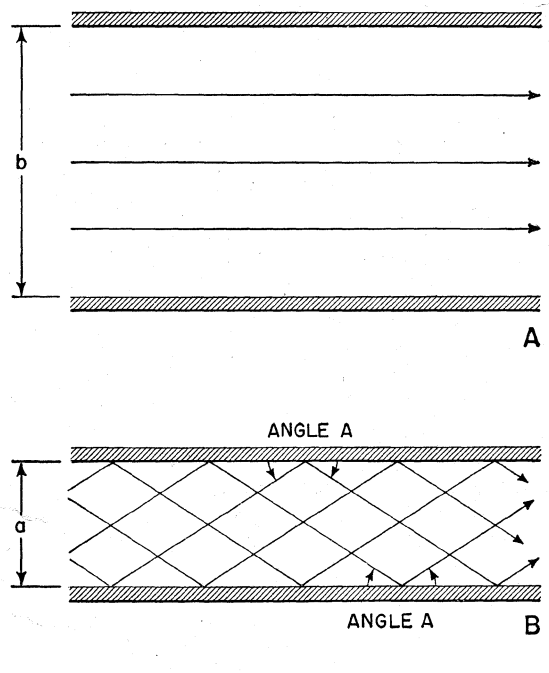


Figure 52. Paths of two-wave patterns.

f. Behavior of Tipped Patterns. Each of the patterns formed by combining two sets of waves has a $\lambda_a/\cos B$, where the wavelength is measured in a direction parallel to the path of the pattern. The pattern moves along the path (fig. 52) with a velocity $c/\cos B$; energy moves along the path with a velocity $c \cos B$. The wavelength and the two velocities are equal to the guide wavelength and the phase and group velocities of the $TE_{0,n}$ mode. As might be expected, the patterns cannot exist unless λ_a is shorter than $2b/n$, the cut-off wavelength for the mode. The two patterns combine in the same way as the two sets of waves that form a $TE_{m,0}$ mode, and angle A must satisfy the equation

$$\sin A = m\,\lambda/2a,$$

where λ is the wavelength of each of the patterns, or $\lambda_a/\cos B$.

g. Field Configuration of Complete Pattern. The general appearance of the combined pattern is shown in figure 53, where the direction of the electric field is shown by solid lines and that of the magnetic field is shown by dashed lines. The pattern for vertically polarized waves is indicated in A, and the pattern for horizontally polarized waves is shown in B. In each pattern, both m and n are equal to 1. The pattern for higher values of m and n is simply a repetition of the simple pattern, either up or down or across the guide, as in figures 50 and 51. To simplify the

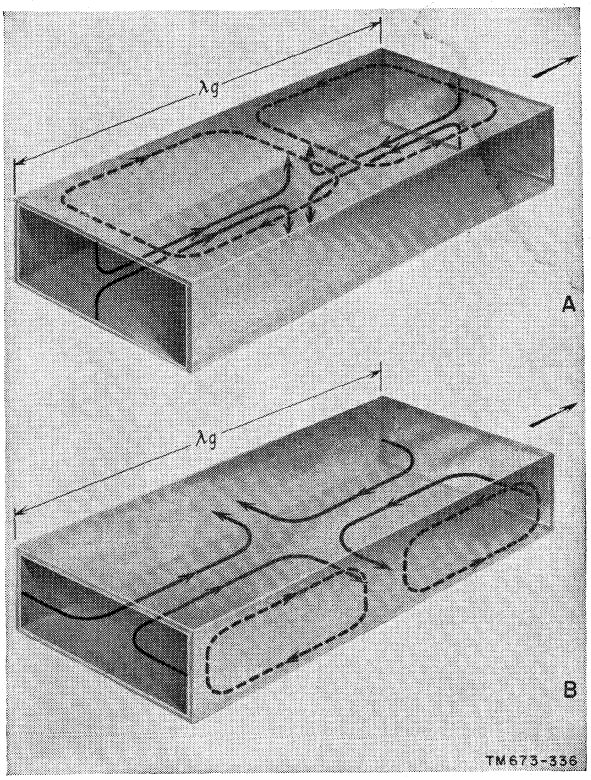


Figure 53. Configuration of combined pattern.

diagrams, most of the lines that show field direction have been omitted. The direction of the electric field is shown in a plane through the center of the guide, a vertical plane for the vertically polarized waves and a horizontal plane for the horizontally polarized waves. Similarly, the direction of the magnetic field is shown only near the top of the waveguide and near the right-hand wall. The pattern of the magnetic field near the bottom of the waveguide is the same as the pattern near the top, with the directions reversed. A similar situation, with identical patterns and reversed directions, exists for the magnetic fields near the two side walls.

h. $TE_{m,n}$ and $TM_{m,n}$ Modes. If eight sets of electromagnetic waves are present in a rectangular waveguide, both of the patterns shown in figure 53 will be present. If the phase relation between the two patterns is the same as that in the drawing, the longitudinal component of the magnetic field will be canceled and the result will be a $TM_{m,n}$ mode. If, on the other hand, the phase of one of the patterns is reversed (by reversing the phases of the four sets of waves that contribute to it), the longitudinal electric field will be canceled, and a $TE_{m,n}$ mode will result. In figure 54, A and B show the direction of the resulting electric and magnetic fields in a plane across the waveguide at a point where the transverse field is at a maximum.

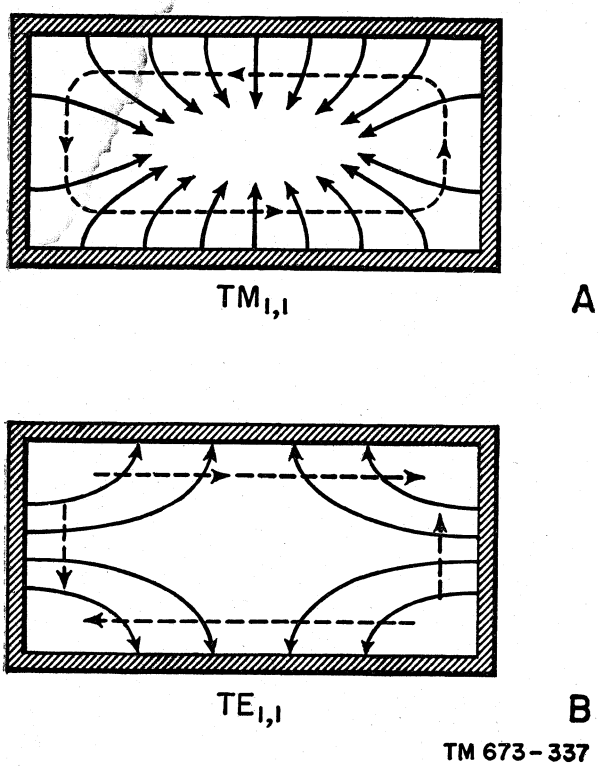


Figure 54. $TM_{1,1}$ and $TE_{1,1}$ modes.

i. Single-Mode Operation of Rectangular Waveguide. Assume that the dimensions of a rectangular waveguide are chosen so that λ_c is shorter than λ_a for the $TE_{1,0}$ mode. The waveguide then cannot carry energy at λ_a by means of the $TE_{1,0}$ mode, and this mode is said to be *suppressed* for λ_a and the corresponding frequency. The cut-off wavelength becomes shorter as either m or n is made larger. Consequently *all* modes for which m is 1 or a larger number will be suppressed. (It is not necessary to specify the values of n because *any* value is equal to zero or a larger number.) The only possible modes are, therefore, those for which m is zero. Since the $TM_{0,n}$ modes do not exist, these are the $TE_{0,n}$ modes. If, in addition, the $TE_{0,2}$ mode is suppressed, all modes for which n is 2 or greater also will be suppressed. There remains only the $TE_{0,1}$ mode (since neither the $TE_{0,0}$ nor the $TM_{0,0}$ mode exists). This procedure normally is followed with rectangular waveguide. Height a is made less than $\lambda_a/2$ so that the $TE_{1,0}$ mode, which has a cut-off wavelength of $2a$, is suppressed. The $TE_{0,2}$ mode, which has a cut-off wavelength of b, is suppressed by making b less than λ_a. The cut-off wavelength for the $TE_{0,1}$ mode is $2b$, and if the width is more than $\lambda_a/2$ (but less than λ_a), the waveguide will operate in this mode only.

j. Dimensions of Rectangular Waveguide. It is desirable to use waveguide as large as possible, provided that undesired modes are suppressed. Large waveguide has less attenuation than small waveguide, just as a large wire has smaller losses (because of its lower resistance) than a small wire when both carry the same current. Modes other than the $TE_{0,1}$ mode are suppressed for much the same reason that standing waves on a transmission line are suppressed. Both absorb power, and both interfere with the proper operation of the system by changing impedance. It is usual to use 1½- by 3-inch rectangular waveguide for S-band operation and ½- by 1-inch rectangular waveguide for X-band operation. Some of the characteristics of these two sizes of rectangular waveguide are tabulated below. In practice, the shortest operating wavelength is 5 or 10 percent longer than the wavelength at which double mode operation is possible. This insures that the waveguide will operate only in the $TE_{0,1}$ mode even when there is a small error in the operating frequency. The longest operating wavelength is about 80 percent of λ_c for the

$TE_{0,1}$ mode. The waveguide will operate at longer wavelengths (so long as they are shorter than λ_c) but the attenuation increases rapidly as the operating wavelength approaches the cut-off wavelength. A good rule of thumb for estimating the operating wavelength of an odd size of rectangular waveguide is that the nominal operating wavelength is about five-fourths of the larger dimension.

	Nominal operating wavelength (S-band)[a]	(X-band)[b]
Outside dimensions in inches	1.500 x 3.000	.500 x 1.000
Wall thickness in inches	.080	.050
Inside dimensions in inches	1.340 x 2.840	.450 x .950
Inside dimensions in centimeters	3.40 x 7.22	1.14 x 2.41
λ_c, $TE_{0,1}$ mode	14.44 cm	4.82 cm
λ_c, $TE_{0,2}$ mode	7.22 cm	2.41 cm
λ_c, $TE_{1,0}$ mode	8.80 cm	2.28 cm

[a] S-band—10 cm.
[b] X-band—3 cm.

56. Circular Waveguides

The behavior of circular waveguide is similar to that of rectangular waveguide. The field patterns, like those of rectangular waveguide, satisfy the following two *boundary conditions* imposed by the fact that the waveguide wall is a good conductor: At the surface of a wall there can be no electric field parallel to the wall, and there can be no varying magnetic field perpendicular to the wall. A review of the material on rectangular waveguides will show that these boundary conditions are satisfied by the field patterns of all existing modes. Because of the curved walls of circular waveguide, it is extremely difficult to examine the field patterns in the same detail as those of rectangular waveguide—there is no simple method of *building* the field pattern from a small number of simple waves traveling along the guide. The calculation of cut-off wavelength for a given mode in circular waveguide is, therefore, beyond the scope of this manual, and it will be necessary to accept the values given here without any demonstration of their validity. The general configuration of the field patterns may be deduced, however, from their similarity to corresponding patterns in square waveguide (rectangular waveguide for which a and b are equal).

a. Terminology. As with rectangular waveguide, all circular waveguide modes are TE modes, TM modes, or combinations of these, and TE and TM have the same significance as they have in rectangular waveguide modes. The modes are described further by the same two subscripts, so that there are $TE_{m,n}$ modes and $TM_{m,n}$ modes. In rectangular waveguide, m is the number of times that the field pattern is repeated from the top of the guide to the bottom, with a reversal in the directions of the electric and magnetic fields at each repetition, and n is the number of times that the pattern is repeated from one side wall to the other. In circular waveguide, however, both m and n have different meanings: m is the number of times the pattern is repeated *around* the waveguide, and n is the number of times the pattern is repeated *between the center of the guide and the wall.*

b. Important Modes. In considering the behavior of circular waveguide, it is necessary to examine only the $TE_{1,1}$, $TM_{0,1}$, and $TE_{2,1}$ modes. These are the three simplest modes for a circular waveguide and the only ones for which the cut-off wavelength is greater than the diameter of the guide. $TE_{1,1}$ is the *dominant mode* (that having the longest cut-off wavelength) for a circular waveguide; it has a cut-off wavelength of $1.71d$, where d is the inside diameter of the waveguide. In figure 55, A shows the electric field pattern for the $TE_{1,1}$ mode, and B shows the electric field for the $TE_{0,1}$ mode in a square waveguide (a rectangular guide for which a and b are equal). The resemblance between the fields in the circular and square waveguides is obvious, and both modes are dominant modes. The $TM_{0,1}$ mode, with a cut-off wavelength of $1.31d$, is shown in A, figure 56; the $TM_{1,1}$ mode in a square waveguide is shown in B. The two diagrams are simplified by omitting the lines showing the direction of the magnetic field. The $TE_{2,1}$ mode with a cut-off wavelength of $1.03d$,

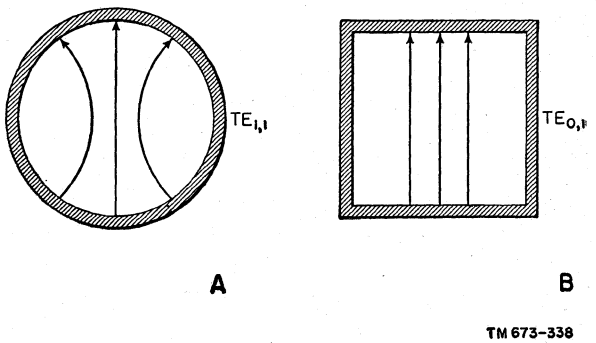


Figure 55. Dominant TE modes in circular and square waveguides.

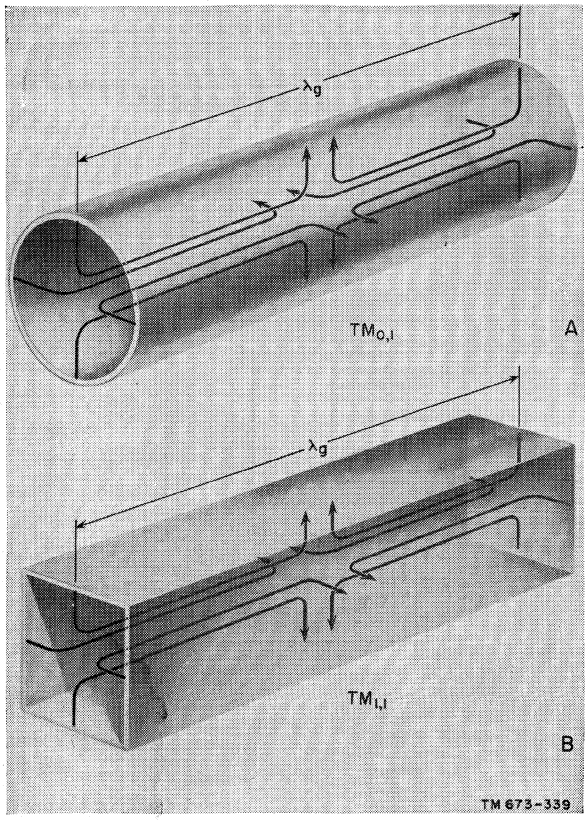


Figure 56. Simplest TM modes in circular and square waveguides.

is shown in A, figure 57. The corresponding mode for a square waveguide is the $TE_{1,1}$ mode, shown in B. All other modes, both *TE* and *TM*, have cut-off wavelengths of .82*d* or less.

c. Usefulness. The usefulness of circular waveguide lies in the perfect symmetry of the $TM_{0,1}$ mode. A rotating joint can be inserted in a section of circular waveguide without affecting the transmission of energy. Therefore, a section of circular waveguide is used to join two sections of rectangular waveguide when it is necessary to move one in relation to the other (fig. 58). Moving the upper rectangular waveguide from the position shown in A to the position shown in B has no effect on the transmission of energy. Although it is possible to use the $TE_{1,1}$ mode of circular waveguide, it is difficult with this mode to control the plane of polarization of the signals, and the corresponding $TE_{0,1}$ mode in rectangular waveguide normally is used instead.

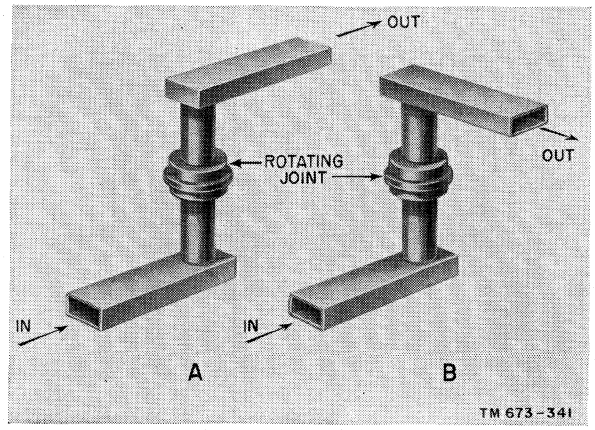


Figure 58. Rotating joint in circular waveguide.

d. Dimensions. Circular waveguide always is chosen so that λ_a is shorter than 1.31*d*, the cut-off wavelength for the $TM_{0,1}$ mode. The higher *TM* modes then are automatically suppressed by choosing the diameter so that λ_a is longer than .82*d*. The method of exciting the waveguide suppresses the $TE_{1,1}$ and $TE_{2,1}$ modes unless the second of these already is suppressed by a choice of diameter. It is usual to excite the guide with a small antenna at the guide axis (fig. 59). The field radiated by this antenna has no longitudinal magnetic component (which must be present for

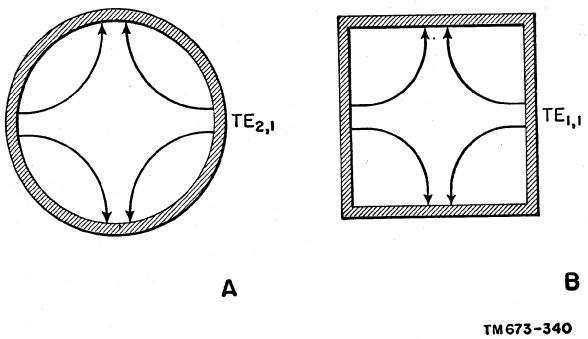


Figure 57. Simplest TE modes in circular and square waveguides.

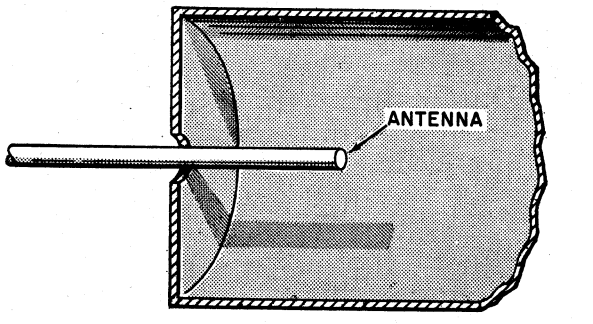


Figure 59. Excitation of circular waveguide.

any *TE* mode), and, consequently, neither the $TE_{1,1}$ nor the $TE_{2,1}$ mode is excited. Mechanically, it is desirable to use circular waveguide with a diameter equal to the width of the rectangular waveguide in the system, so that d is normally about $.8\lambda_a$, but this is not always done. The circular waveguide normally used in S-band systems has a diameter of about 3 inches, and that used in X-band systems is about 1 inch in diameter.

57. Waveguide Advantages

The major advantage of waveguide over coaxial line is the much lower loss of waveguide, resulting primarily from the elimination of the center conductor. At vhf and microwave frequencies, interaction between the guided waves in a transmission system and the current-carrying electrons causes the current flow to be confined to the surface of the conductors. This *skin effect*, of course, is present at lower frequencies, but it is unimportant. However, at vhf and microwave frequencies, because of the skin effect, only the surface of the inner conductor of coaxial line is available for current carrying and its resistance is unavoidably high. The outer conductor, having a larger surface (the inner surface is the one that carries the current) has a much lower resistance. The major portion of the loss, therefore, occurs at the inner conductor of coaxial line, which is not present in waveguide, and it is not present when waveguide is used. Waveguide has other advantages, such as lack of radiation and dielectric losses, but rigid coaxial line offers the same advantages. The greatest merit of waveguide, aside from the fact that it has lower conductor loss (sometimes called copper loss) than coaxial line, is its extreme simplicity of construction. By comparison with waveguide, a rigid coaxial line with quarter-wave stub supports is a complicated mechanical device.

Section III. TUNING DEVICES AND RELATED EQUIPMENT

58. Components Used with Coaxial Line

Before considering the accessories used with waveguide, it is desirable to examine some of those used with coaxial line at microwave frequencies because their operating principles are more familiar than those of the waveguide components. The coaxial-line accessories are specialized to a certain extent because of the construction of the line itself. In general, they are rigid structures which, like the coaxial line, avoid the use of any insulating material other than dry air. Several of the more usual components are tuning stubs, slotted sections, line stretchers, and rotating joints.

a. Tuning Stubs. A tuning stub is a short section of coaxial line connected across the main line. The arrangement is similar to that of the quarter-wave stub support, except that the length of the tuning stub can be adjusted to any value between zero and a half-wave length. Two tuning stubs usually are mounted together on a single section of coaxial line inserted in the main transmission line, the spacing between the two stubs usually being three-eighths wave length (fig. 60). The inner and outer conductors have the same dimensions as those of the remainder of the coaxial line, and a single quarter-wave stub support holds the inner conductor of the section in its proper place. A flange, provided at each end of the section, makes contact with the outer conductor of the adjacent portion of the line. The inner conductor has a socket at one end and a finger at the other end; these parts connect with matching parts on adjacent sections of the line and provide continuity of the inner conductor. The finger usually is split and slightly spread apart to make a good friction fit in the matching socket. The two stubs are identical. Each consists of an inner conductor, an outer conductor, and an adjustable plunger which acts as a short circuit. The length of the stub is varied by changing the position of the plunger. A number of small metal brushes are provided on the plunger to insure good contact with both the inner and the outer conductors. The plunger is carried on one end of a hollow screw which has an adjusting knob at the other end; the screw itself is carried in a threaded cap mounted on the end of the outer conductor. A pressure seal with packing under it serves to prevent loss of air from the coaxial line and also acts as a lock for the adjusting screw. There are, of course, many different ways of constructing a double-stub section, but the one shown in figure 60 is a good illustration of the principles involved in the types in general use.

b. Stub Tuning Theory. Every transmission line has a characteristic impedance, usually written as Z_o, which depends on the physical characteristics of the line. If a current I flows in the line, the

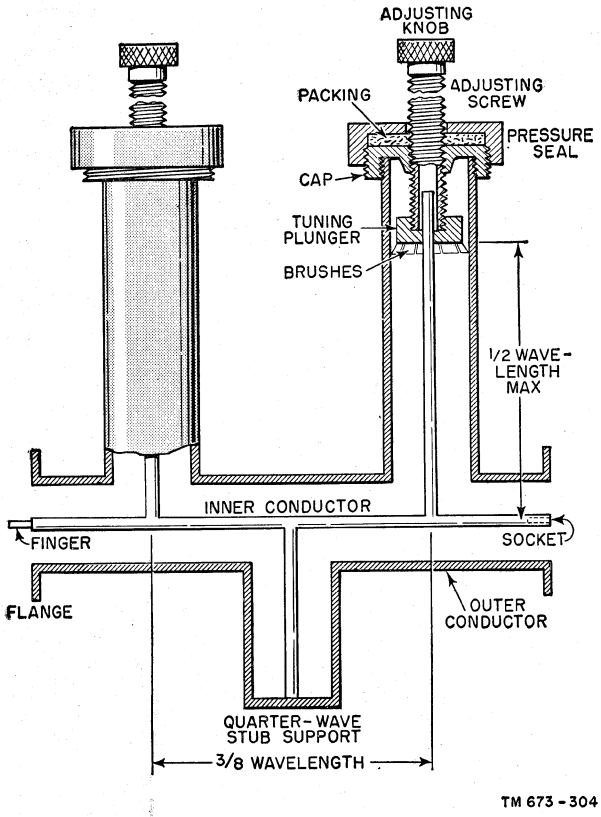


Figure 60. Double-stub section of coaxial line.

voltage E accompanying it then is derived from the equation

$$E=Z_oI$$

This equation holds for signals traveling in either direction along the line, so that

$$E_1=Z_oI_1$$

and, by division,

$$I_1=E_1/Z_o,$$

and

$$E_2=Z_oI_2, \text{ or } I_2=E_2/Z_o$$

where E_1 and I_1 are the voltage and current carrying power from a generator to a load, and E_2 and I_2 are the voltage and current of the reflected signal carrying power in the opposite direction. Figure 61 shows the positive directions for E_1, E_2, I_1, and I_2. For convenience, the transmission line is shown as a two wire line since the discussion applies to this type as well as to a coaxial line. It can be seen that the voltage measured at any point will be the sum (E_1+E_2) and the current will be the difference (I_1-I_2) when the phase relations between E_1 and E_2, and between I_1 and I_2 are taken into account. If, now, the transmission line is terminated by a load of impedance Z, as in figure 62, and if E and I are the voltage and the current at the load, E/I must be equal to Z. In this case,

$$Z=E/I=(E_1+E_2)/(I_1-I_2)$$

But, since $I_1=E_1/Z_o$, and $I_2=E_2/Z_o$, these values can be substituted in (I_1-I_2),
and the equation

$$E_2=E_1(Z-Z_o)/(Z+Z_o)$$

is obtained by simple algebra. The method of working the equation to its conclusion is shown in appendix II. If Z is equal to Z_o, $(Z-Z_o)$ will be equal to zero, and E_2 (and consequently I_2) also will be equal to zero. In simpler terms, there is no reflected signal when a transmission line is terminated in its characteristic impedance. The reflected signal is objectionable for two reasons, the first of which has to do with power losses on the transmission line. Suppose that the loss on a given transmission line is 10 percent. An input of 100 watts then will provide 90 watts at the load if there is no reflected signal. If, however, there is a reflected signal, more input power is required to deliver the same 90 watts to the load. Suppose, for example, that E_2 is one-half of E_1. I_2 also will be one-half of I_1, and the power in the reflected signal will be one-fourth of the power in the direct signal. If, now, 120 watts arrives at the load end of the transmission line, 30 watts will be reflected and

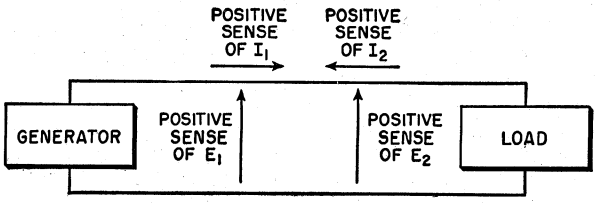


Figure 61. Directions of direct and reflected signals in transmission line.

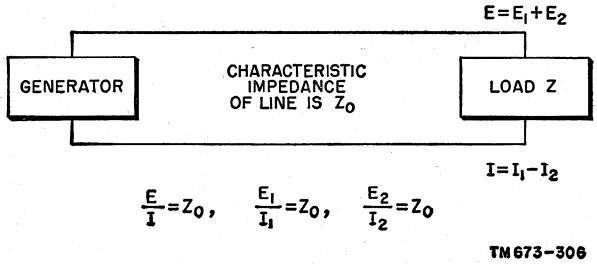


Figure 62. Generation of reflected signal.

90 watts will be delivered to the load. In order to deliver 120 watts at the load end, however, it is necessary to put 133⅓ watts in at the generator end of the line because of the 10-percent transmission line loss. This extra power requirement is offset partially by the reflected signal, which delivers 27 watts (because of the 10-percent transmission line loss) at the generator end. The generator itself then is called on to supply 106⅔ watts (the difference between 133⅓ watts and 27 watts) in order to develop 90 watts at the load. This is an increase of 6⅔ watts because of the reflected signal, and any power increase is undesirable. The second objection to the reflected signal is the possibility of an impedance mismatch at the generator. If Z is equal to Z_o, there is no reflected signal, and the voltage and current at the input end of the transmission line will have the ratio equal to Z_o. The generator customarily is designed to deliver maximum power output when it sees this impedance. In the presence of a reflected signal, however, the impedance seen by the generator is equal to the ratio $(E_1+E_2)/(I_1-I_2)$. This ratio is, of course, equal to Z at the load end. It is necessary, however, to take account of phase when adding and subtracting the voltages and currents. For a position between the generator and the load, the phases of E_1 and I_1 are advanced, and those of E_2 and I_2 are retarded. Because of this, the impedances measured at different points on the line have different values and, in particular, the impedance measured at the generator may have values as far apart as Z and Z_o^2/Z. With proper matching, the generator in the example above may develop 200 watts, of which 100 watts is delivered to the transmission line, and the rest is lost in the generator. When there is a reflected signal of the strength given above, the generator is required to supply 106⅔ watts to the transmission line. An impedance mismatch resulting from a change in the input impedance of the line, however, may require that the generator develop 400 watts in order to transfer 106⅔ watts to the line. The tuning stubs shown in figure 63 are used when the load impedance is not equal to the characteristic impedance of the line. They are placed on the line close to the load, and are adjusted so that the reflected signals from the stubs cancel the reflected signal from the load. The portion of line between the stubs and the generator is, therefore, free of reflected signal. Each stub delivers a reflected signal because it is a section of line terminated in a zero impedance.

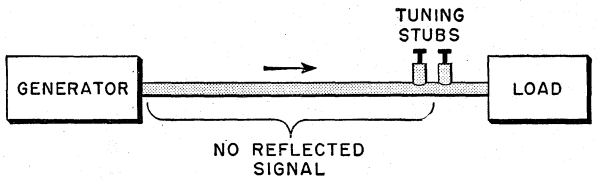


Figure 63. Use of tuning stubs to cancel reflected signals.

c. Tuning Stub Adjustment. In practice, adjustment of the tuning stubs is started by setting both stubs at their midpositions. Each is then a quarter-wavelength and looks like an infinite impedance across the line (so that the stubs have no effect). The stub nearest the generator then is moved, while the load power is observed. An initial adjustment is reached for which the load power is a maximum. In a radar, this may be accomplished by observing a fixed echo and adjusting the stub for maximum echo strength. After the initial adjustment, the second stub, nearer the load, is moved a little way and the first stub is readjusted for maximum power. Depending on whether this is an improvement, the second stub then is moved farther in the same direction or back in the opposite direction. The process continues until the best adjustment has been found by trial, always with a readjustment of the first stub after the second one has been moved.

d. Slotted Section. When it is not convenient to observe the load power while the tuning stubs are being adjusted, it is necessary to make use of a slotted section of coaxial line (fig. 64) so that the voltage on the transmission line may be measured at various points. The measured voltage is the sum of E_1 and E_2, where these are added vectorially because of the phase difference between them. At some point, the two voltages will be in phase and the measured voltage will have a maximum value. At another point, the two voltages will be 180° out of phase and the measured voltage will have a minimum value. Then, if $E_{1\text{-peak}}$ is the peak value of E_1, and $E_{2\text{-peak}}$ is the peak value of E_2, with E_{max} and E_{min} the maximum and minimum values of the measured voltage,

$$E_{max}=E_{1\text{-peak}}+E_{2\text{-peak}},$$

$$E_{min}=E_{1\text{ peak}}-E_{2\text{-peak}}.$$

The ratio E_{max}/E_{min} is called the *VSWR* (*voltage standing wave ratio*) and it is a measure of the amount of reflected signal that is present. If the

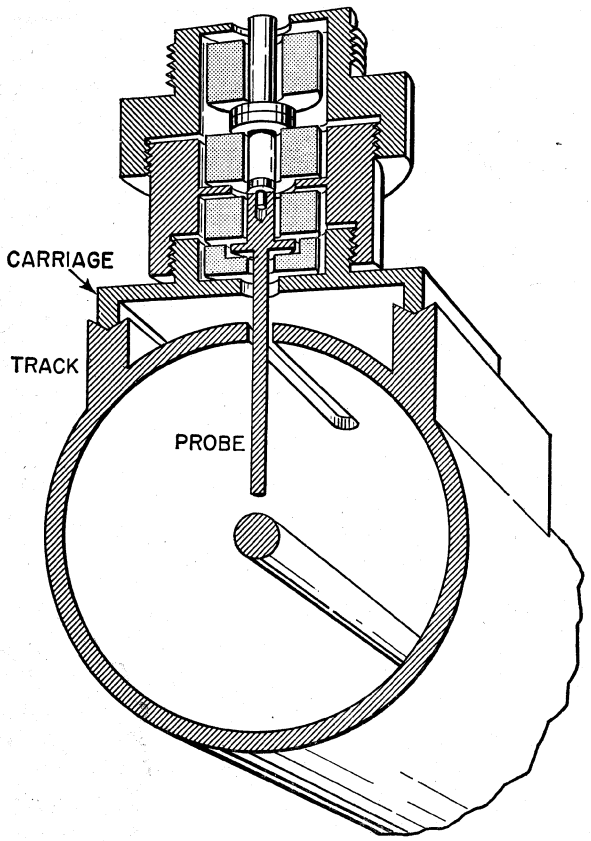


Figure 64. Slotted section of coaxial line.

VSWR is 1, there is no reflected wave. The adjustment of the tuning stubs then may be accomplished by observing the VSWR instead of the power in the load, and adjusting for the lowest VSWR. Values of 3 usually are considered acceptable, in operation at microwave frequencies, although lower values are desirable if they can be obtained. The slotted section is simply a section of line with a slot in the outer conductor so that the inner conductor becomes accessible. A probe is mounted on a small carriage, which moves on a track fixed to the outer conductor. The probe does *not* touch the inner conductor, and a signal is picked up by virtue of the capacity between the inner conductor and the tip of the probe. It is important that the spacing between the probe and the center conductor remain constant as the probe is moved, and the carriage and track must be inspected from time to time. The signal picked up by the probe is rectified and sent to an indicator of some sort, usually a milliammeter which is preceded by an amplifier. The VSWR is found by moving the carriage back and forth and noting the maximum and minimum indicator readings, the ratio of these being the VSWR. The slotted section may be quite short. If the two voltages add to produce a maximum at one point, they will do so at other points that are spaced an integral number of half-wavelengths away. This occurs because a position change of 1 half-wavelength shifts the phase of E_1 by 1 half-cycle in one direction and the phase of E_2 by 1 half-cycle in the other direction, thus leaving them in the same phase relation as before. Midway between each two adjacent points where the measured voltage is a maximum, there is a point where E_1 and E_2 are 180° out of phase and the measured voltage is a minimum. It is always possible, then, to find one maximum and one minimum if the slot is 1 half-wavelength long. In practice, the slot is longer than this, and several readings are taken at different maxima and minima. A slotted section frequency is built in as part of a transmission line. If, however, the line is to be pressurized, it is necessary to provide a removable section of ordi-

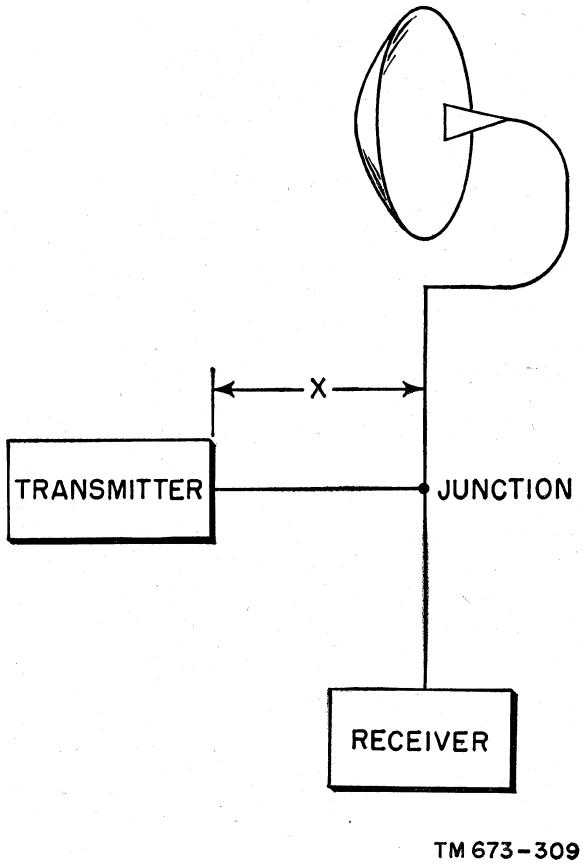


Figure 65. Installation with precision transmission line length.

nary coaxial line which can be replaced by the slotted section when the VSWR is to be measured.

e. Line Stretcher. In some installations, a line is terminated by an impedance different from Z_o and the length of the line is adjusted so that the input impedance has a specified value (fig. 65). In this example, if the length, X, is adjusted properly, the impedance looking toward the transmitter from the junction will be infinite when the transmitter is turned off (between pulses). The transmitter, therefore, does not take any energy from the received signal. Because of frequency variations in different transmitters, it is not practical to adjust length X at the factory. Instead, a device called a line stretcher (fig. 66) is installed between the transmitter and the junction as part of the transmission system. Its operation is obvious, since it consists of a U-shaped section which slides up and down to change the length of the line. It is the coaxial equivalent of a slide trombone.

f. Rotating Joint. It is frequently necessary for one part of a transmission line to move while another part remains fixed. When the system consists of rigid coaxial line, this freedom is obtained by using a rotating joint. A of figure 67 shows a quarter-wave section of coaxial line terminated by an open circuit. This section has zero input impedance; in other words, it acts like a short circuit at its input end. In practice, some capacity always exists at the terminating end and the section must be made slightly less than a quarter-wave length. In B, the same section is shown with the inner conductor made hollow. It is then possible to bring a pair of conductors, P and Q, to the sending end of the quarter-wave section and provide a good electrical connection between them without any mechanical contact. In figure 68, B shows this device used to make a rotating joint in coaxial line, and A shows a practical joint developed from the scheme used in B. Such a joint is used frequently, but it is difficult to seal because of the high voltage between M and N. It is, therefore, modified to the form shown in figure 69 when the system is pressurized. It is possible to make a seal between U and V without difficulty, since these points can be connected without affecting the electrical operation. The resistance of the seal is in series with the infinite input impedance of the shorted quarter-wave section to the right of V, and, whether the seal is good or bad from an electrical standpoint, the quarter-wave section to the left of U always is terminated in the

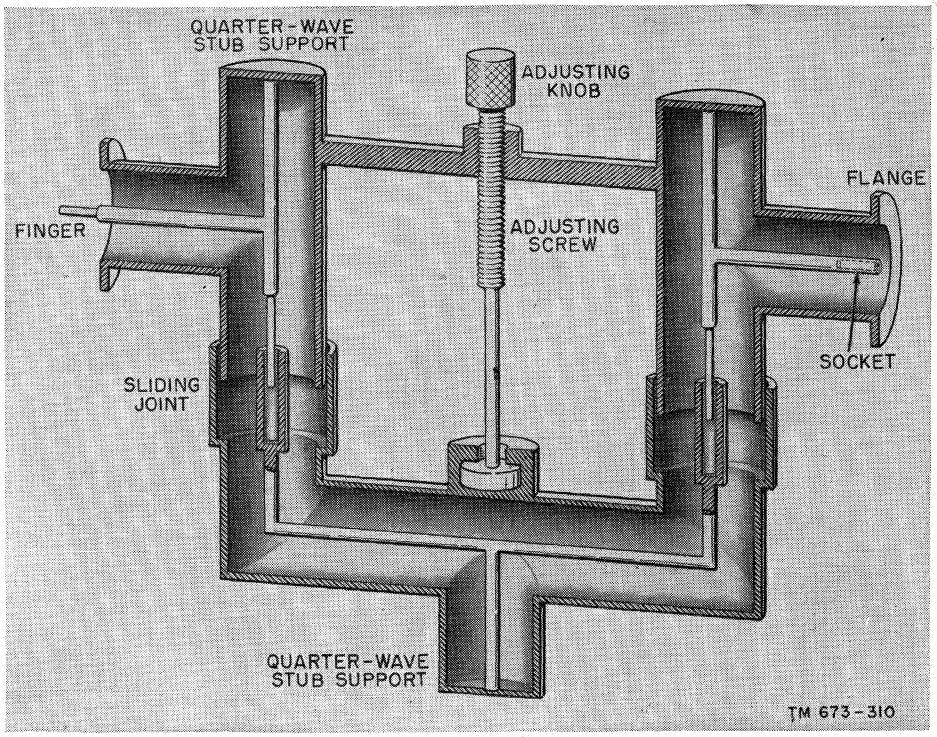


Figure 66. Coaxial-line stretcher.

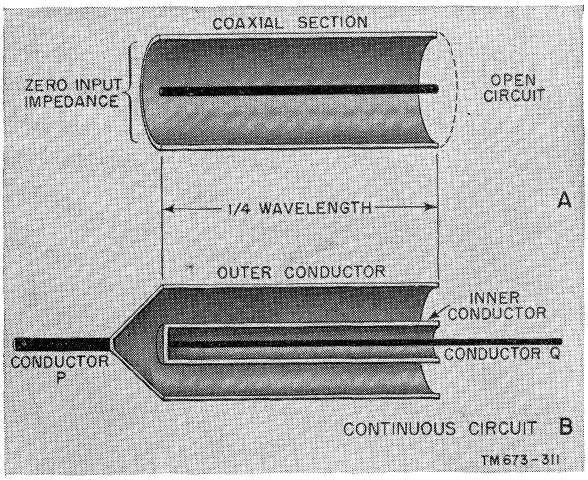


Figure 67. Use of quarter-wave coaxial section for electrical connection with mechanical freedom.

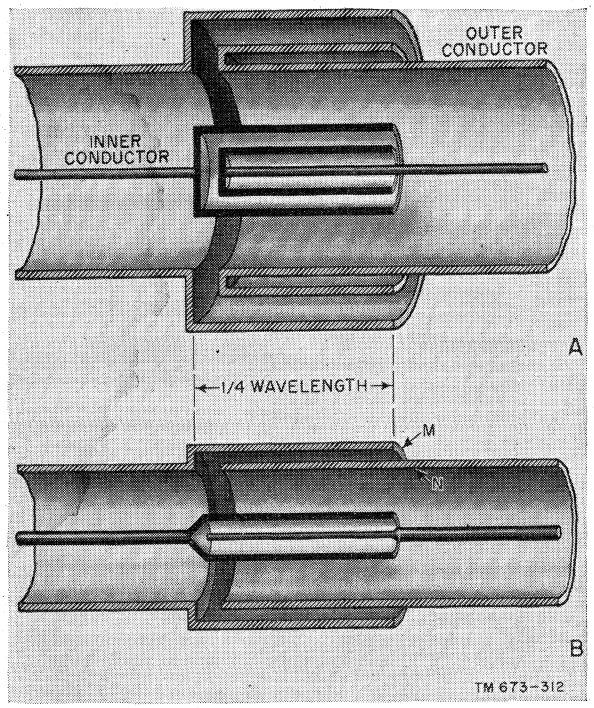


Figure 68. Rotating joints.

required open circuit. By use of a half-wave section, the connection for the outer conductor then is terminated in a short circuit, which also has zero input impedance.

59. Waveguide Tuning Devices

It seems inappropriate, at first glance, to talk about the impedance of a waveguide, since there is no point at which *the* voltage and *the* current can be measured. The concept of wave-

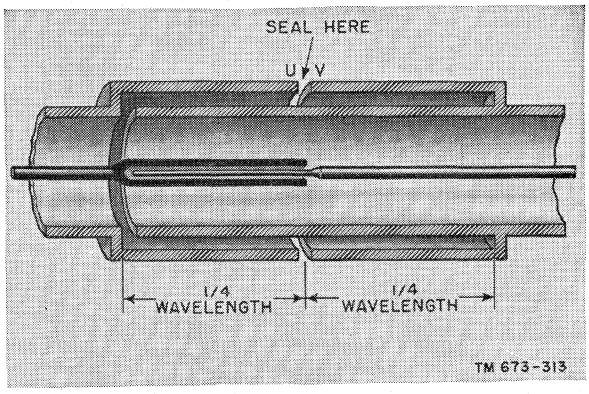


Figure 69. Pressurized rotating joints.

guide impedance is, however, as useful as that of transmission line impedance.

a. Waveguide Impedance. The impedance of a transmission line at any point is the ratio between the voltage and the current at that point. It is necessary, of course, when computing this ratio, to take into account the phase relationship between the voltage and the current. The impedance of a waveguide at any point is, similarly, the ratio between the transverse electric field and the transverse magnetic field at that point. As with transmission line impedance, it is necessary, when computing the ratio, to take into account the phase relationship between the two fields. A, of figure 70 shows the fields in a coaxial line, and B shows the fields in a conventional two-wire line. In both, the transmission line acts as a guide for the guided waves that accompany any transmission of energy. This was mentioned briefly in paragraph 45 and was discussed in some detail, for a coaxial line, in paragraph 47. At any point where the guided wave is present, such as X, in A, and Y, in B, electric and magnetic fields are associated with the guided waves. The transverse component of the electric field (the field in a plane perpendicular to the conductors) depends on the position of the point and on the voltage between the conductors. For any given point, the strength of the electric field is proportional to the voltage. In the same way, the transverse magnetic field depends on the position of the point and on the current carried by the conductors. For any given point, the strength of the magnetic field is proportional to the current. Both the transverse electric field and the transverse magnetic field depend in the same way on the position of the point, so that the ratio between them does not depend on the position of the point.

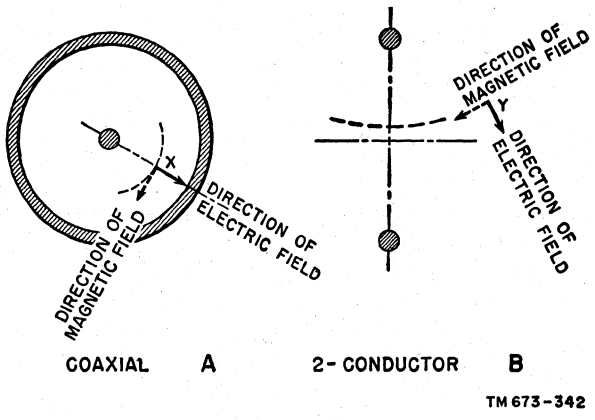


Figure 70. Fields associated with transmission lines.

The ratio between the two transverse fields is, then, proportional to the ratio between the voltage and the current and it follows that the impedance of a transmission line can be defined either by the ratio between the voltage and the current or by the ratio between the transverse electric field and the transverse magnetic field. The second way of defining the impedance can be applied directly to a hollow waveguide. The waveguide, then, has a characteristic impedance associated with each mode (although several modes may have the same characteristic impedance). As explained in paragraph 48 for the coaxial line, the actual impedance is equal to the characteristic impedance when no reflected wave is present. In the presence of a reflected wave, the impedance is different from the characteristic impedance and is different at different points along the waveguide. The impedance is always the same at all points in a plane perpendicular to the axis of the waveguide, just as the impedance of a transmission line (computed from the electric and magnetic fields instead of from voltage and current) is the same, whatever point in a given plane is chosen as a place to measure the fields.

b. Slug Tuner. A slug tuner is used with a waveguide transmission system in precisely the same way that the stub tuner described in paragraph 58 is used with a conventional transmission line, and the slug tuner is essentially the same device as the stub tuner. Most waveguide systems consist largely of rectangular waveguide operating in the $TE_{0,1}$ mode, and the double-slug tuner shown in figure 71 is designed for this type of waveguide. A shows a section through the tuner, and B shows an end view. Each tuning slug projects into the waveguide parallel to the electric field and acts as a small antenna. It is excited by the electric field in the waveguide, and, consequently, radiates a signal. The phase of the reflected signal may be altered by changing the length of the slug with the tuning screw, and the signals reflected from the slugs cancel the signal reflected from the load. Electrically, the double-slug tuner is similar to the double-stub tuner shown in figure 60, with one exception: The range over which the slugs can be varied is less than the range over which the stubs can be varied. The tuning stub has no effect when it is a quarter-wavelength, and its length may be varied in either direction from this point. The slug is ineffective when it is backed out to a position even with the waveguide wall, and it can be varied in only one direction from this point. Consequently, the double-slug tuner is limited to some extent, and three-slug tuners frequently are used with spacing of 3/8 λ_g between the adjacent slugs. A double-slug tuner is adjusted by starting with both slugs retracted so that their tips are even with the wall of the waveguide, the position in which they have no effect. One of the slugs then is inserted by advancing the tuning screw. If this makes the situation worse, that slug is retracted and the other is advanced. From this point, tuning proceeds by alternate adjustment of the two slugs. This method is similar to the one described in paragraph 58 for the double-stub tuner.

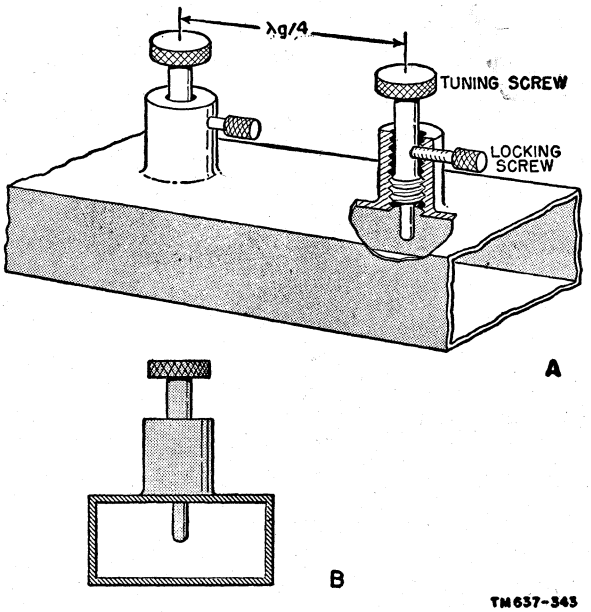


Figure 71. Double-slug tuner for $TE_{0,1}$ mode in rectangular waveguide.

c. Tuning Stubs. Although actual tuning stubs are used occasionally, in most applications they offer little advantage over the slug tuner. A typical tuning stub is shown in figure 72. It consists of a section of waveguide with an adjustable plunger of conducting material. To insure good contact between the plunger and the wall of the waveguide and, therefore, an effective short circuit for the termination, the plunger is grooved a short distance back from its face. The groove, when seen from B, looks like a short section of transmission line terminated by the short circuit at A. The groove depth is chosen so that this section of line has an effective length of 1 quarter-wavelength. The impedance at B, therefore, is the infinite impedance of a shorted quarter-wave section of line, and this is in series with the impedance of the rubbing contact between the plunger and the waveguide wall to the left of B. Distance BC is also 1 quarter-wavelength, and the impedance of the gap between the plunger and the wall at C is that of a quarter-wave section terminated by the impedance at B. Since the impedance at B is always great (because of the impedance of the groove), the impedance at C is almost zero, *whether or not there is a good contact between the plunger and the wall.* The groove in the plunger is called a *choke groove.* The stub may be added to either wall of the waveguide. When the junction is at the wall perpendicular to the electric field, it is called an *E-junction,* the stub is called a *series stub,* and the impedance of the stub is in *series* with that of the main line. When the junction is at the wall perpendicular to the magnetic field, it is called an *H-junction,* the stub is called *shunt stub,* and the impedance of the stub is in *parallel* with that of the main line. A of figure 73 shows a stub connected with an E-junction, and, in B, the stub is connected with an H-junction. In practice, the E-junction is the one normally used.

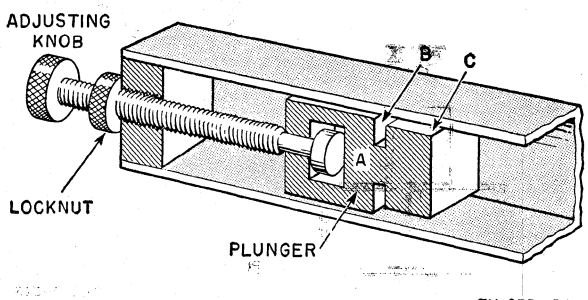


Figure 72. Adjustable tuning stub.

d. Window Tuners. It is always possible to tune a waveguide by any method that introduces the required reflection, and this sometimes is done by means of window tuners (fig. 74). Since it is desirable for the two obstructions to be symmetric, it is rather difficult to construct an adjustable window tuner. However, window tuners are useful whenever they can be preset and left alone. Electrically, they are equivalent to shunt tuning stubs. The window tuners are nothing more than small conducting plates inserted in the guide. The tuner shown in A is located where the electric field is relatively weak, and its major effect is on the magnetic field. Since the magnetic field is analogous to current in the impedance of a waveguide, a tuner of this sort behaves like a current device and is approximately the equivalent of a shunt inductor (which does not affect the voltage) on a conventional transmission line. The tuner shown in B affects the electric field more strongly than the magnetic field and is the approximate equivalent of a shunt capacitor.

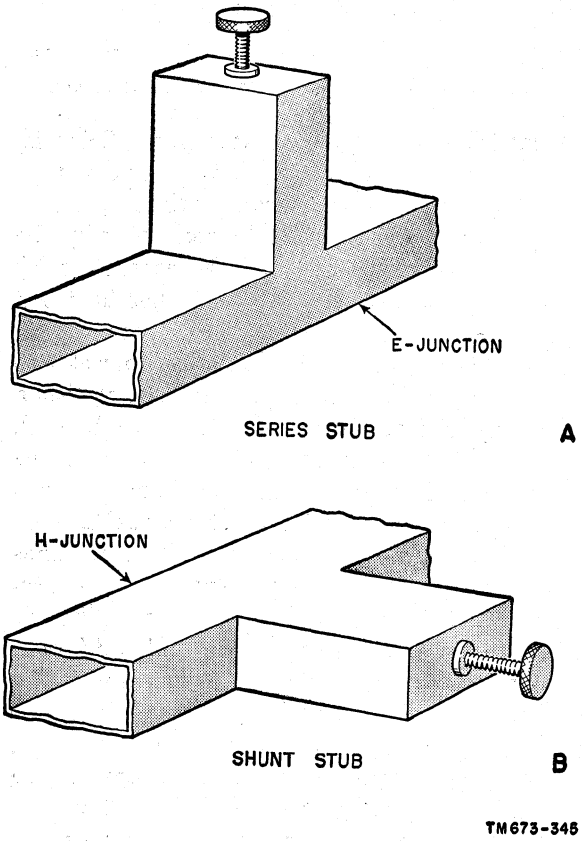


Figure 73. Shunt and series tuning stubs.

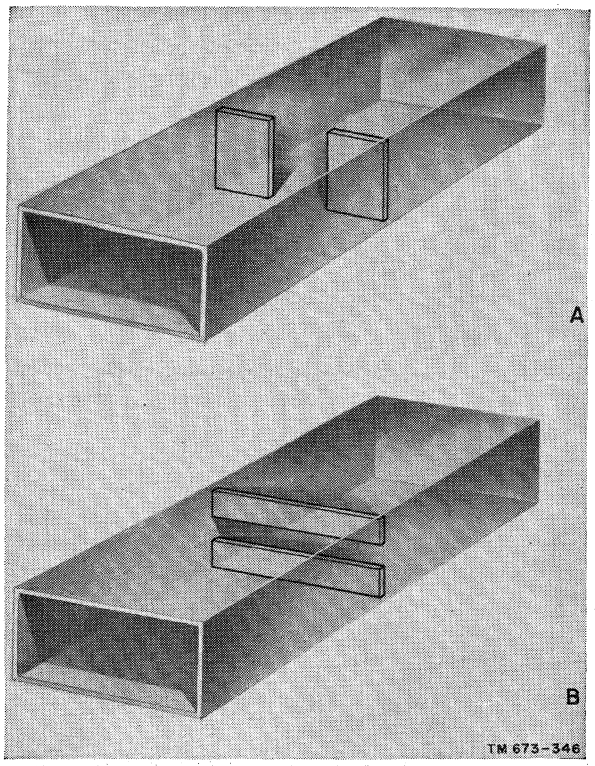


Figure 74. Window tuners

60. Slotted Sections

Whenever a waveguide carries both direct and reflected signals, the strength of the electric field varies from point to point along the guide, just as voltage varies in a conventional transmission line. A slotted section with coaxial line is used to observe the load power, as it is used with waveguide (par. 58*d*). Following transmission line terminology, the ratio between the maximum and the minimum electric field strengths is called the VSWR. Figure 75 shows a cross section of the carriage which rides on a slotted section of rectangular waveguide. The probe, which is parallel to the electric field, is excited by the signal and develops a voltage at the operating frequency. A crystal rectifier is mounted in the carriage, and the output of the crystal is fed to a meter. Generally, it is possible to use a sensitive meter directly but in some cases an amplifier is required. Figure 75 shows the carriage and crystal holder with a fitting for a coaxial cable.

61. Directional Couplers

A directional coupler (fig. 76) generally is used to measure the reflected signal in a waveguide. The directional coupler requires no manipulation

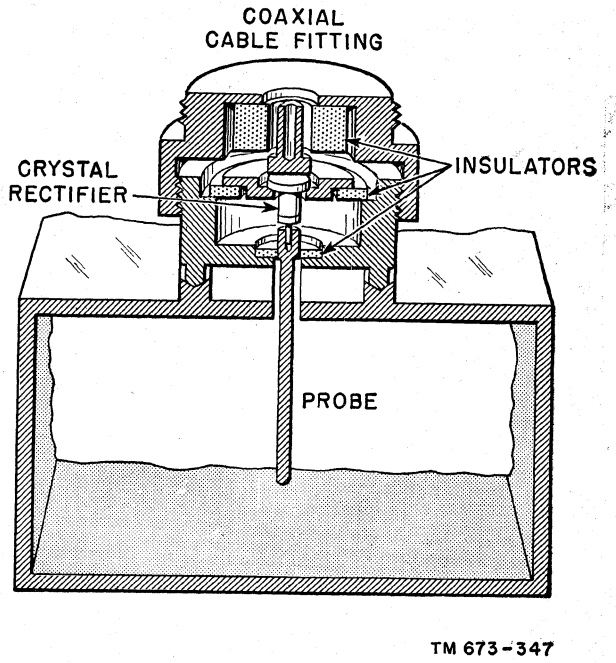


Figure 75. Slotted section.

and can be sealed to make a permanent part of a pressurized waveguide system. Figure 77 shows the basis for the theory underlying the use of the directional coupler. In figure 76, a main waveguide has an auxiliary waveguide coupled to it by means of two small holes. Referring, now, to A of figure 77, it is seen that when a signal travels from left to right in the main waveguide, some energy leaks into the auxiliary guide through the holes at WX and YZ, and that the signal in the auxiliary guide at Z is composed of two parts. Some energy leaks through the hole at WX and travels from X to Z. Additional energy travels from W to Y and then leaks through the hole at YZ. Since distance WY and XZ are equal, the two parts of the signal will be in phase and some energy will be propagated from left to right in the auxiliary guide. Propagation of a signal in the opposite direction in the auxiliary guide, as illustrated in B, depends on the spacing between the holes. Some energy leaks directly through the hole at WX. Additional energy travels from W to Y, leaks through the hole at YZ, and then travels back from Z to X. The longer path differs from the shorter path by twice the spacing between the holes. If, then, the holes are separated by a distance $\lambda_g/4$, the two signals will be out of phase at X and *no energy will be propagated from right to left in the auxiliary guide.* Energy from the direct signal leaks into the directional coupler

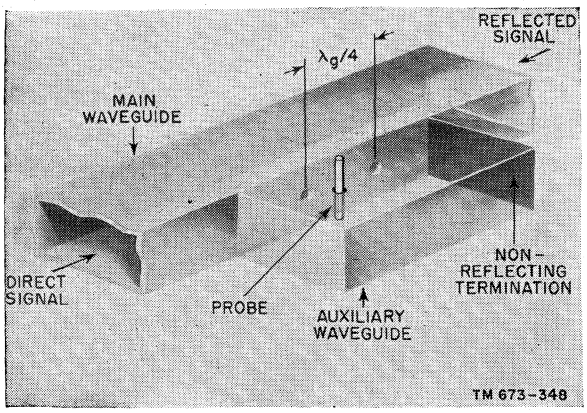


Figure 76. Directional coupler.

and travels in the same direction to the right-hand end, where it is absorbed by a nonreflecting termination which is a resistive surface, the impedance of which matches that of the waveguide (fig. 76). Energy from the reflected signal leaks into the directional coupler and continues to the left-hand end, where it excites the probe. Any reflection from the left-hand termination of the directional coupler reaches the right-hand termination and is absorbed. The probe is fitted with a crystal holder, as is the carriage of the slotted section shown in figure 75. The signal picked up by the probe is proportional to the strength of the reflected signal in the main waveguide and does not depend on the level of the signal traveling in the other direction. A waveguide system frequently is fitted with two directional couplers, one to measure the strength of the desired signal and one to measure the level of the reflected signal. It is possible also to feed energy *into* the directional coupler through the probe. Some of this energy will be absorbed by the resistive termination, and the rest will leak through the two holes into the main waveguide. Because of the difference in the path lengths through the two holes, the energy that reaches the main waveguide will be propagated *in only one direction.*

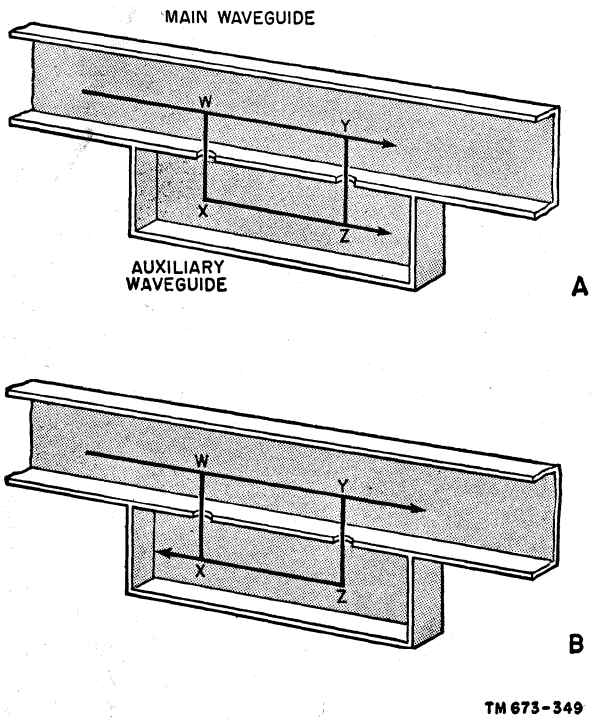


Figure 77. Theory of directional coupler.

62. Attenuators

It is sometimes necessary to adjust the power level in a waveguide, and this is done by means of an attenuator. Any of the various tuning devices will serve as attenuators, since they can be adjusted deliberately for an impedance mismatch. There then will be a reflected signal, and only a portion of the incident signal in the waveguide will be transmitted, since the remainder is reflected. A more convenient device, however, is the shutter attenuator shown in A of figure 78.

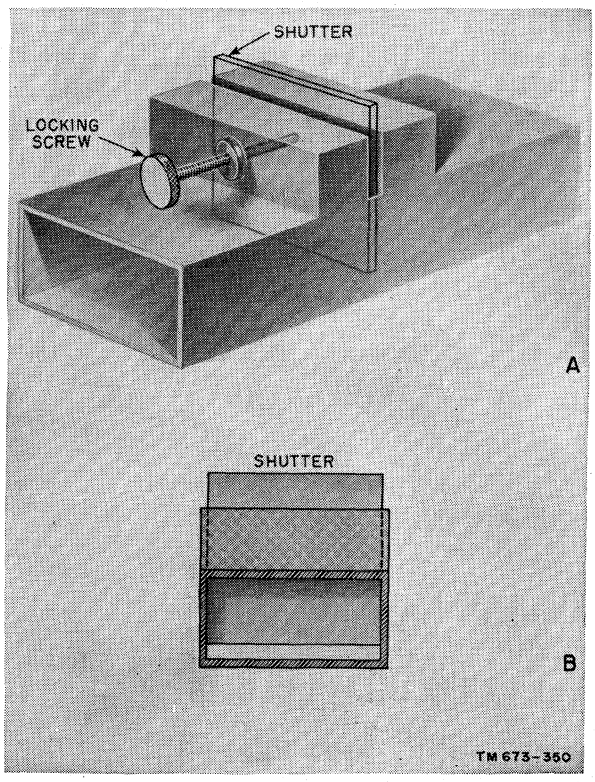


Figure 78. Shutter attenuator.

B shows how the conducting shutter closes off most of the waveguide so that only a small amount of energy is passed. In many cases, it is desirable to attenuate the signal without introducing any reflection, and this is accomplished by means of a dissipative attenuator (A of fig. 79).

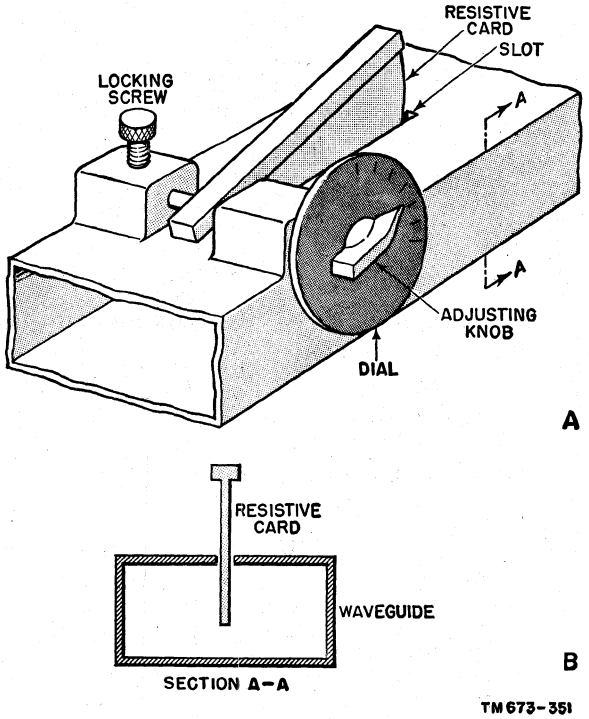


Figure 79. Dissipative attenuator.

The structure consists of a holder carrying a card coated with a resistive material. This card is inserted in the waveguide, as in B, parallel to the electric field and it absorbs energy. If the resistance of the card has the proper value, the dissipative attenuator will not introduce reflection. The resistance of a conducting surface of this sort is specified in *ohms per square* without any unit of distance. In figure 80, A shows a square of conducting surface which may have any convenient dimensions—1 inch by 1 inch, for example. Two conducting bars are fastened to opposite sides of the square, and the resistance between these is taken as R ohms. In B, the rectangle is 1 inch by 2 inches, with a resistance of $2R$ ohms. However, two of these rectangles in parallel form the larger square, 2 inches by 2 inches, as shown in C, and this square has a resistance of R ohms. It follows that the resistance of any square will be the same as the resistance of any other square, and the behavior of the material is specified completely if the ohms per square is known. The impedance of a transmission line is specified at some point, and that of a waveguide is specified at some plane perpendicular to the waveguide axis. The waveguide is, therefore, like a conducting surface and its impedance is given in ohms per square.

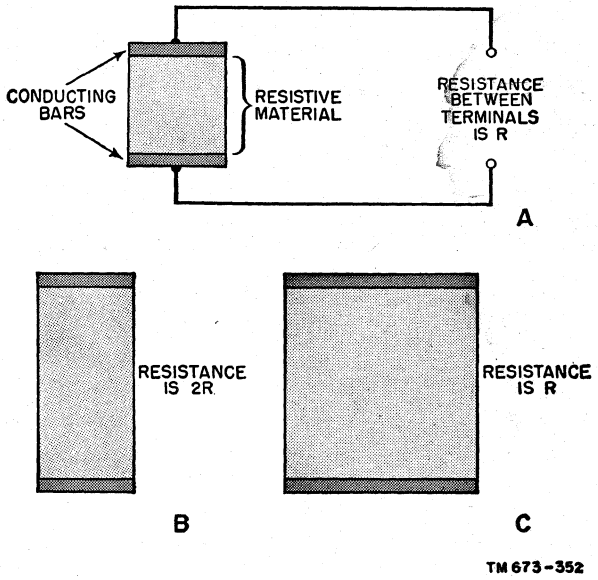


Figure 80. Resistance of conductive surface.

Section IV. COUPLINGS, JOINTS, BENDS, AND TERMINATIONS

63. Coupling from Transmitter to Transmission System

A microwave transmitter consists of a simple oscillator (together with its power supplies, modulator, and control circuits) operating at the desired output power level. Because of the difficulty of making a power amplifier for microwave frequencies, a power oscillator is used instead of the oscillator and power amplifier normally employed at lower frequencies. The oscillator is normally a multicavity magnetron, although a reflex klystron is used frequently in low-power applications (ch. 5). In either case, microwave power is developed in a resonant cavity, or cavity resonator (ch. 4) which is an integral part of the tube structure. Energy is taken from the cavity on a short length of coaxial line, which is also a part of the tube structure, and delivered to the transmission

system, which may be either coaxial line or waveguide.

a. Magnetron Coupling. In many S-band magnetrons the coaxial line from the resonant cavity is terminated in fittings that allow it to be joined directly to a standard S-band coaxial transmission line. Figure 81 is a photograph of such a magnetron. In magnetrons used at higher frequencies, and sometimes in S-band magnetrons, the coaxial line from the resonant cavity is coupled to a short section of waveguide that is also a part of the magnetron structure. This, then, is connected to the waveguide transmission system which carries the output of the magnetron. Although either rectangular or circular waveguide *may* be used for the output fitting, most magnetrons use rectangular waveguide. Figure 82 shows a magnetron of this sort.

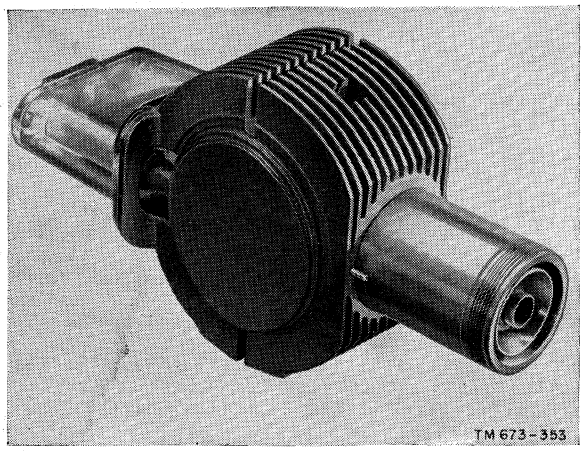


Figure 81. S-band magnetron with coaxial line output fittings.

b. Impedance Matching Between Magnetron and Line. A magnetron is designed to have an output impedance equal to the characteristic impedance of the transmission line with which it is used. (*Transmission line* as used here, means either a coaxial line or a hollow waveguide.) The magnetron will, therefore, operate with reasonable efficiency when it is connected directly to a transmission line of the type it is designed to feed. For highest efficiency, however, it is usually necessary to insert a tuning device between the magnetron and the transmission line. The tuning device is either a stub tuner (par. 58*a*) or a slug tuner (par. 59*b*), depending on whether the transmission line is coaxial line or waveguide. As stated in the discussion on the theory of tuning devices of this sort (par. 58*b*), reflected signals are

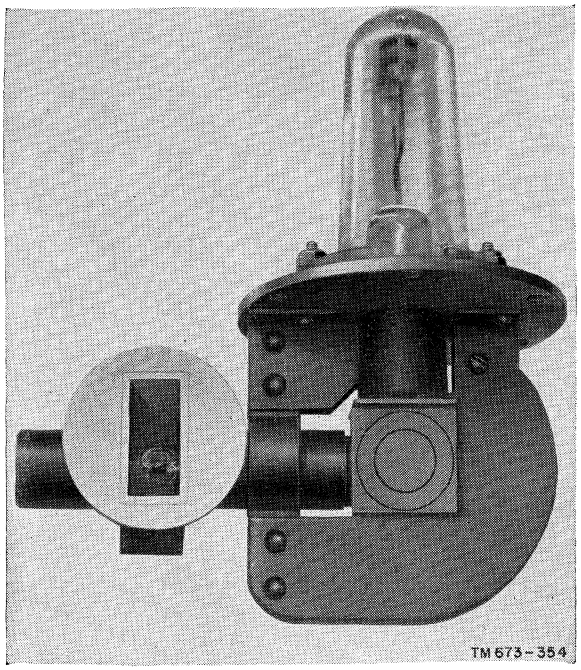


Figure 82. Magnetron with waveguide output fitting.

introduced into the coaxial line or waveguide and the input impedance of the tuner is made to take any desired value. When the tuner is adjusted properly, the magnetron delivers maximum power to the transmission line.

c. Transmission Line Tuning. A complete transmission system, coupling a magnetron to a load, is shown in figure 83. Tuner X matches the transmission line to the magnetron, and tuner Y matches the load to the line. The VSWR is measured at the slotted section (normally used with a coaxial-line system), and tuner Y is adjusted until the VSWR is as close as possible to unity. With this adjustment, there is no reflected signal to the left of tuner Y and the load is matched to the transmission line. The line to the left of tuner Y is said to be *flat* when there is no reflected signal. Following adjustment of tuner Y, tuner X is adjusted so that the magnetron delivers maximum power to the line, the signal level being measured at the slotted section. With a waveguide transmission system, the slotted section generally is replaced by two directional couplers, although the slotted section can be used with waveguide. One directional coupler reads the reflected signal, and tuner Y is adjusted so that this has the smallest possible output. The other directional coupler reads the direct signal *from* the magnetron, and tuner X is adjusted so

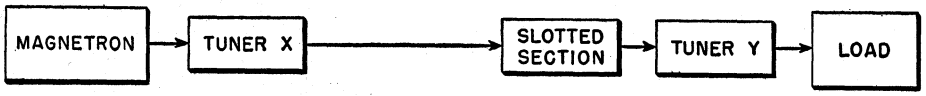


Figure 83. Transmission line with tuners.

that this has the greatest possible output. Tuner Y always is adjusted first.

d. Preplumbed Systems. In many cases, the load and the magnetron are designed to match the transmission line, and both tuners are omitted. More frequently, the tuner next to the magnetron is omitted. The portion of the system from which the tuner is omitted is said to *preplumbed.* In general, preplumbed systems are used when maintenance is difficult to carry out. Although a system with tuners is slightly superior to a preplumbed system when the tuners are adjusted properly, it is inferior when the tuners are adjusted badly. Because of this, the tuners are omitted whenever there is a likelihood that they will get out of adjustment between maintenance checks.

e. Coupling From Klystron to Line. The output of an S-bank klystron normally appears at a coaxial fitting. A conventional polyethylene dielectric flexible coaxial cable (which is kept as short as possible) is used to carry the signal to a rigid S-band coaxial line. When a klystron of this sort is used to feed a waveguide, the coaxial *cable* is connected to a probe of the type described in paragraph 47. Some S-band klystrons, and almost all klystrons operating at higher frequencies, have a short length of rigid small-diameter polystyrene dielectric coaxial line (fig. 28). When used as a transmitter, this type of klystron normally is coupled to rectangular waveguide as shown in figure 84. Although the stub shown at the left may be preplumbed in a transmitter, it is usually adjustable.

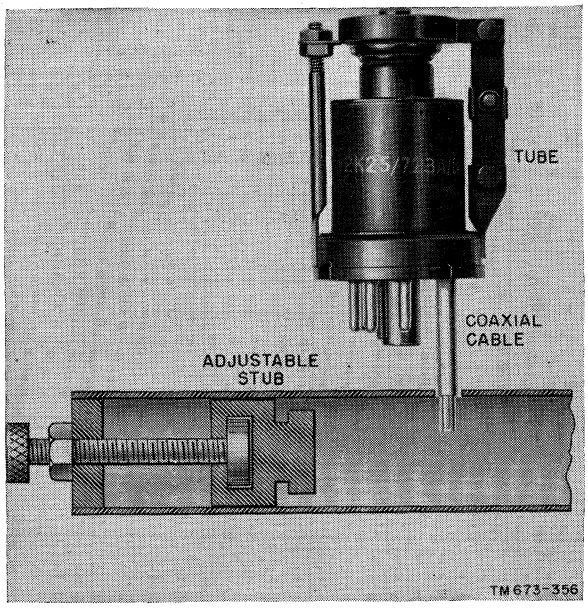


Figure 84. X-band klystron coupled to waveguide.

64. Coupling from Transmission System to Receiver

Although microwave receivers are discussed in detail in chapter 6, the problem of feeding energy from the transmission system to the receiver necessarily is touched on briefly at this point. It is extremely difficult to amplify signals at microwave frequencies and, therefore, a microwave receiver is almost always a superheterodyne in which the input is presented directly to the mixer.

a. Coupling From Coaxial Line. Figure 85, shows a coaxial line receiver coupling assembly. The line is terminated at the crystal mixer on the left, and the i-f output is taken out on a conventional flexible coaxial cable. The local oscillator signal is brought to the assembly on a second flexible coaxial cable and coupled to the mixer through the capacity between the local oscillator probe and the center conductor of the rigid line. The probe sometimes is adjustable but usually is preplumbed. The crystal mixer impedance normally is not equal to the characteristic impedance of the rigid coaxial line, and a two-stub tuner is inserted between the receiver coupling assembly and the transmission system that feeds it.

b. Coupling From Waveguide. Figure 86 shows a waveguide receiver coupling assembly. This assembly is essentially the same as the coaxial-line assembly except that an adjustable stub is used to terminate the waveguide. The coupling is adjusted by *first* finding the best position for the adjustable stub and *then* setting the tuner so that the complete coupling assembly is matched to the waveguide that feeds it. The local oscillator probe is sometimes adjustable but is usually

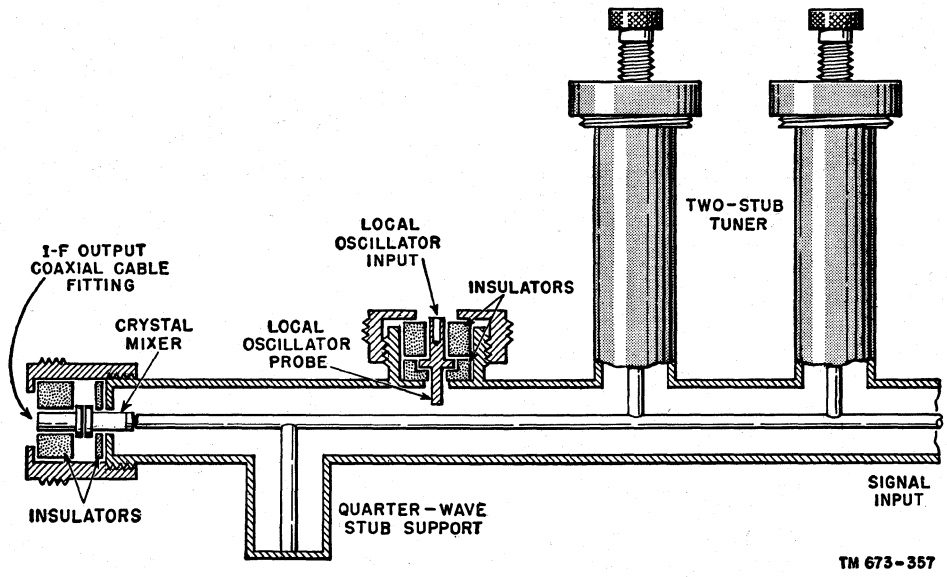


Figure 85. Coaxial line receiver coupling assembly.

preplumbed. The local oscillator signal is injected ahead of the tuner so that both the incoming signal and the local oscillator signal are changed in the same way by adjustment of the stub and tuner. Figure 87 shows the equivalent circuits of coaxial line and waveguide receiver coupling circuits.

c. Low-Sensitivity Couplings. In some cases, it is necessary to measure the level of a fairly strong signal on a microwave transmission line without disturbing the line itself by drawing an appreciable amount of power. This problem arises, for example, with a slotted section. The physical construction of such a coupling is shown in A of figure 88, and the equivalent circuit is shown in B. No tuner is necessary, since the probe projects only a little way into the interior of the coaxial

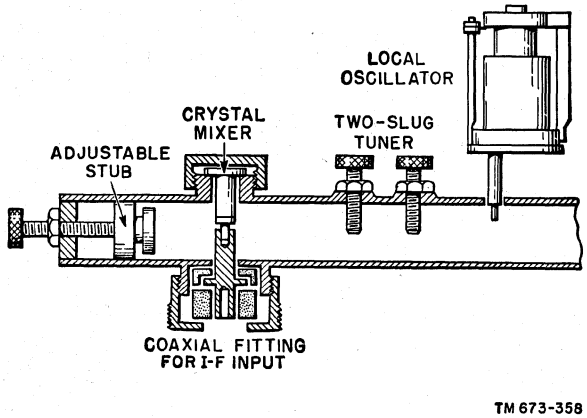


Figure 86. Waveguide receiver coupling assembly.

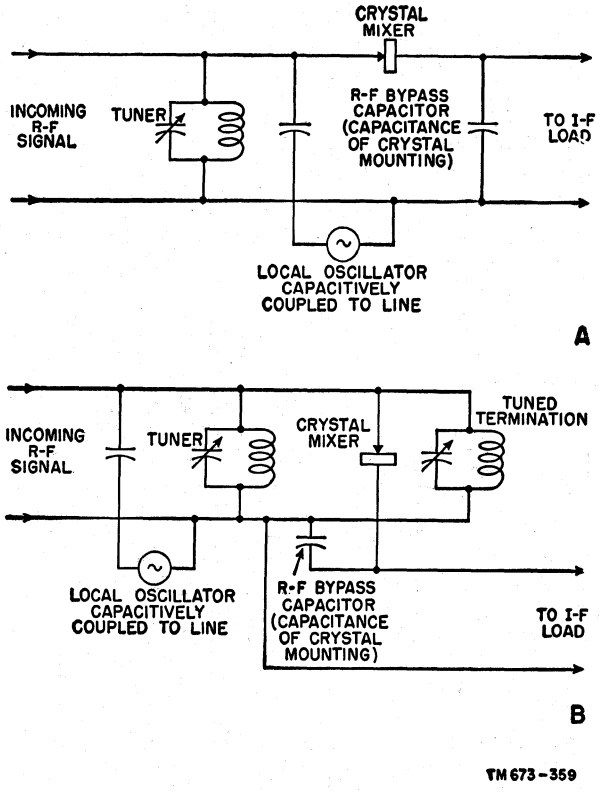


Figure 87. Equivalent circuits of receiver coupling assembly.

line or waveguide. Probe coupling of this type is used with devices measuring the power level directly at the transmission line (as in a slotted section). The output from a directional coupler, however, is taken with the more sensitive arrangement for the crystal (fig. 86), since the

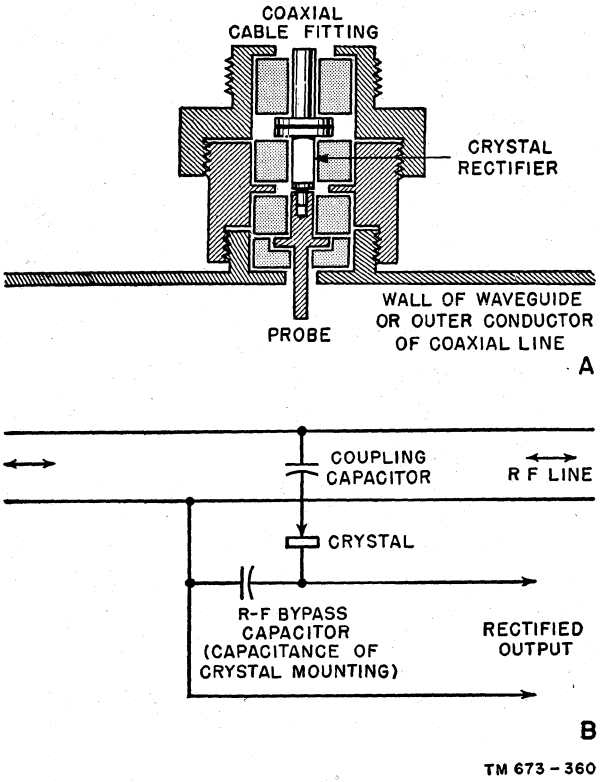


Figure 88. Low-sensitivity coupling and equivalent circuit.

coupler itself is only loosely coupled to the transmission line. (The tuner and the adjustable stub are not used with a directional coupler.)

65. Coupling Between Coaxial Line and Waveguide

Coupling between the miniature coaxial line of an X-band reflex klystron and a rectangular waveguide, where the klystron functions as a transmitter, has been discussed in paragraph 63*e* (fig. 84); a similar coupling is used where the klystron functions as the local oscillator of a receiver (fig. 86). When it is necessary to couple a normal coaxial line to a waveguide, this is accomplished in the manner described below. Rigid coaxial line is shown (fig. 89) but the use of proper fittings will allow flexible coaxial line to be coupled to a waveguide in the same way.

a. Coupling Between Coaxial Line and Circular Waveguide. A coupling of this sort seldom is seen because there is little requirement for it. A shows a coupling that has the merit of being simple but is not very efficient. The coaxial line is terminated at the end wall of the waveguide and the center conductor is extended to form a probe. A more efficient coupling—one in which there is a better impedance match—is shown in B. This is rather bulky, since the tapered section usually is several guide-wavelengths. The coupling shown in B can be improved by the addition of a stub tuner, provided the tuner is adjusted correctly. If there is any possibility of frequency variation under circumstances in which the tuner will not be readjusted immediately, it is preferable to omit the tuner entirely. Both couplings are designed for use with a circular waveguide operating in the $TM_{0,1}$ mode.

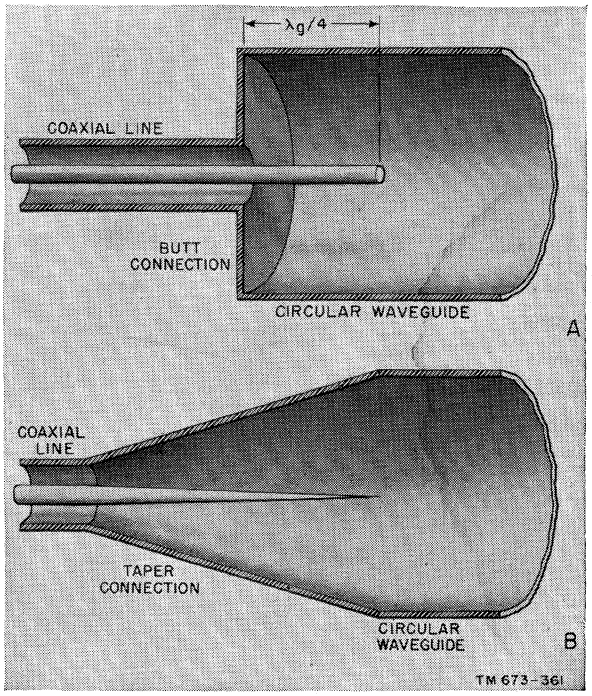


Figure 89. Coupling between coaxial line and circular waveguide.

b. Coupling Between Coaxial Line and Rectangular Waveguide. A of figure 90, shows a simple coupling between coaxial line and rectangular waveguide. This is reasonably efficient, and can be improved by use of either a waveguide or a coaxial line tuner. A more efficient coupling with a preplumbed tuner, shown in B, usually is used with a magnetron fitted with a waveguide output fitting (par. 63*a*). It is possible to eliminate almost all reflection from this coupling by making both the waveguide stub and the coaxial stub adjustable, provided only that the characteristic impedances of the waveguide and the coaxial line are reasonably similar. If the impedances differ

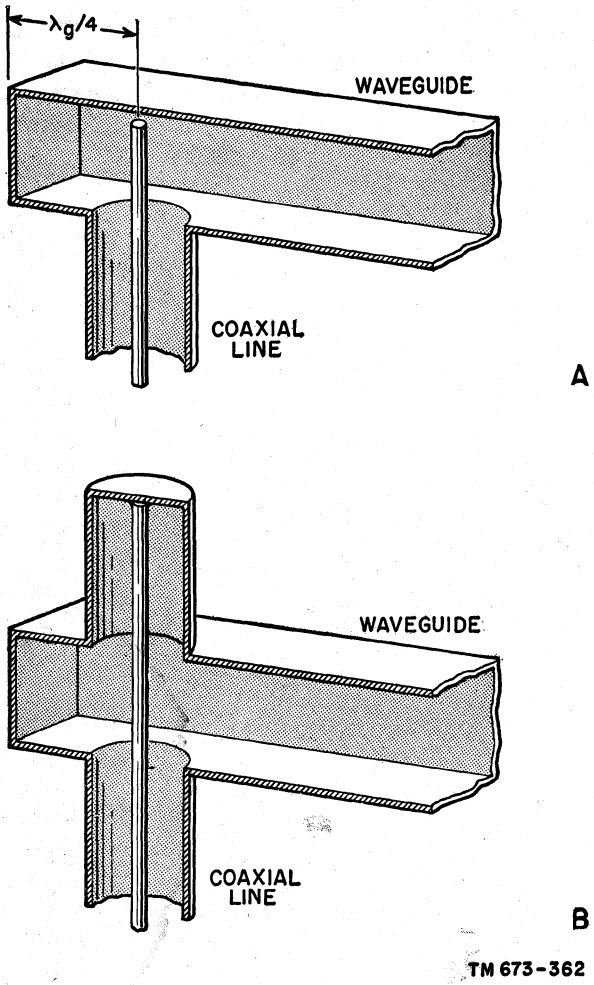


Figure 90. Coupling between coaxial line and rectangular waveguide.

widely, it is sometimes necessary to use a conventional tuner, and in this case the simpler coupling shown in A is used instead. Figure 91 shows a coupling that is used to transfer a small portion of the energy in the waveguide to a coaxial line, or to add a small amount of additional energy to that already present. Such a coupling is used, for example, to feed the signal from a local oscillator into a waveguide connected to the crystal mixer of a microwave receiver, although the arrangements discussed in paragraph 63 are more usual. It is used also to abstract a small amount of energy from the output of a transmitter to feed a wavemeter. The probe extends only a short distance into the waveguide and hardly affects its operation. There is a very poor impedance match, however, at the end of the coaxial line, since it is terminated by a load that is nearly a short circuit. The coaxial line is not tuned because the efficiency is rarely important in applications where this type of coupling is used.

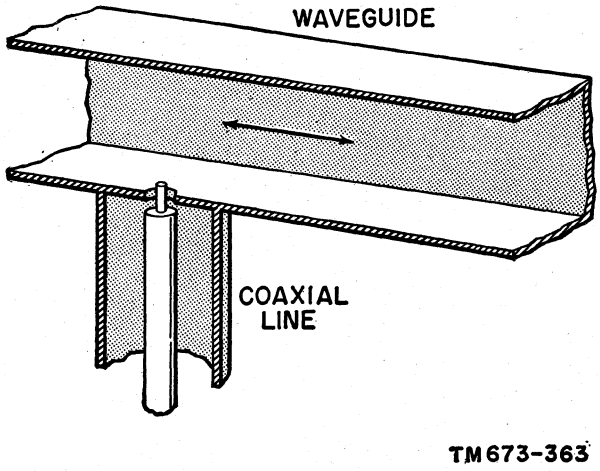


Figure 91. Loose coupling between coaxial line and rectangular waveguide.

66. Coupling Between Rectangular and Circular Waveguides

It is frequently necessary to couple rectangular and circular waveguides together (fig. 92) in order to make use of the rotating joint mentioned in paragraph 58*f*. This type of coupling frequently is preplumbed, but if some adjustment is desired for tuning, the probe is made in the form of a screw so that its length can be varied. For precise matching, the probe is made so that its length is adjustable and the fixed termination at

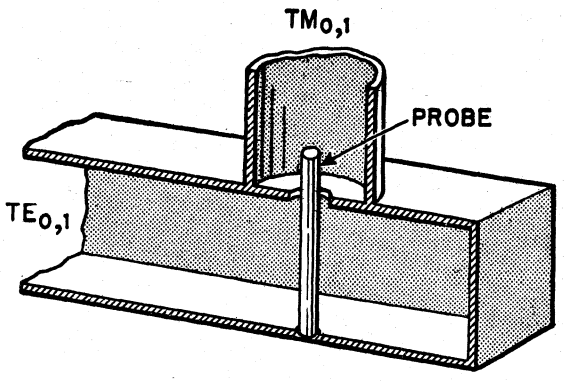


Figure 92. Coupling between rectangular and circular waveguide.

the right is replaced by an adjustable stub, but this refinement ordinarily is not required.

67. Rotating Joints

The rotating joint used with coaxial line already has been described in paragraph 58*f*. The rotating joint used with circular waveguide (fig. 93) consists of two circular plates, one carried on each of the rotating waveguide sections. One of the plates, called a choke plate, is grooved. The other, called a flange, is simply a flat disk. The groove is essentially a shorted section of coaxial line, and its depth is chosen so that it will have an electrical length of a quarter-wave. The impedance at X, looking out from the waveguide, is the impedance of the groove in series with the impedance of the gap between the choke plate and the flange. Since the groove is a quarter-wave line terminated in a short, its impedance is infinite. The total impedance at X is then always infinite for reasonable separation between the choke plate and the flange. Since the mouth of the groove is a quarter-wavelength from the inner wall of the waveguide, the impedance at Y is that of a quarter-wave section terminated by an open circuit, and this is zero. The waveguide, therefore, appears to have continuous walls as long as the two plates are not separated by more than a 1 quarter-wavelength or so. This is not affected by any electrical connection at the circumference of the plates and, consequently, it is possible to seal the joint if the waveguide is to be pressurized. A poor electrical contact at the seal, or even an intermittent one, will not affect the operation of the joint. This system is similar to the one used at the outer conductor of the coaxial line rotating joint. The same method, using a choke groove, is used to insure good electrical termination in an adjustable waveguide stub (par. 59*c*).

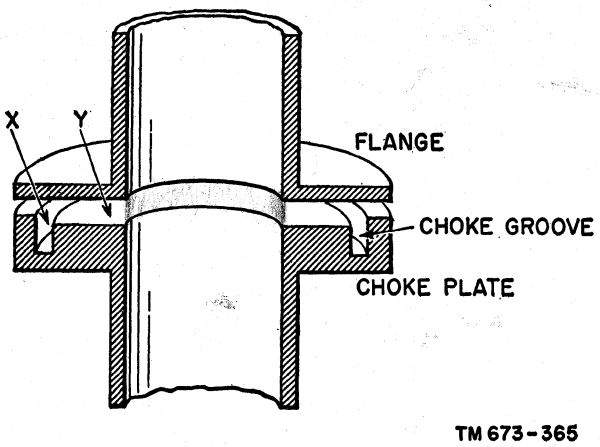


Figure 93. Waveguide rotating joint.

68. Fixed Joints Between Sections of Transmission System

A microwave transmission system normally is made up of a number of sections that can be separated for inspection, cleaning, and possible replacement. Standard joint fittings are used to connect adjacent sections.

a. Fixed Joint in Coaxial Line. Details of the standard S-band coaxial-line joint shown assembled in C of figure 94 with male and female fittings illustrated in A and B, respectively. A clip fastening device holds the two parts of the joint together. Short sections of coaxial line have single quarterwave stub supports to hold the inner conductor in its proper position, and longer sections have stub supports at each end.

b. Fixed Joint in Waveguide. Two sections of waveguide may be joined by providing each one with a flange fitting and securing the joint with several bolts. This normally is not done, however, because undesired reflections are developed at the joint if it is misalined (fig. 95). To minimize the possibility of reflections, a waveguide joint usually is made like the rotating joint shown in figure 93, in which one section is fitted with a choke plate and the other is fitted with a flange. This is called a choke joint and it is preferable to the simple joint in figure 95 because the walls of the waveguide are electrically smooth even though they do not touch each other. Figure 96 shows a cross section of a choke joint. The theory of the choke joint is precisely the same as that of the rotating joint. As explained in paragraph 67, the waveguide walls are electrically continuous at the joint, and the electrical distance between the groove and the inner surface of the waveguide wall is a quarter-wavelength. A of figure 97, shows a choke fitting for use with circular waveguide. A choke fitting for use with rectangular waveguide is shown in B. The distance from the waveguide wall to the groove is, of course, different at different parts of the joint. Distance X–Y, where the electrical field is strongest, is made equal to a quarter-electrical-wavelength, and this is usually satisfactory. A of figure 98, shows an X-band rectangular waveguide with a choke fitting; B is

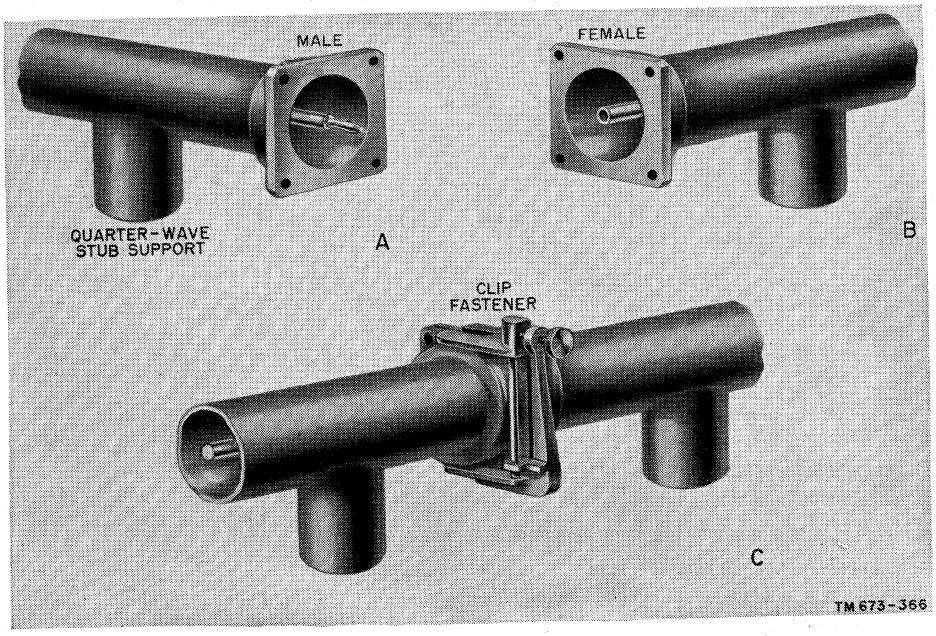


Figure 94. Joint for coaxial line.

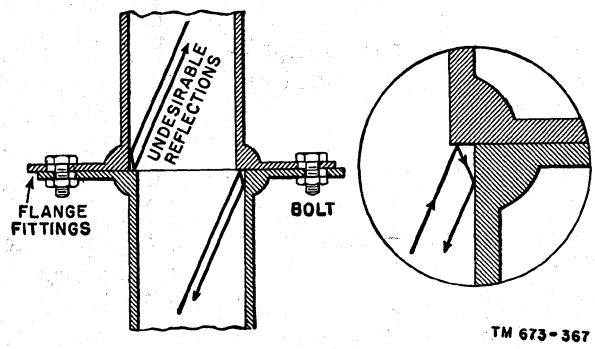


Figure 95. Effect of misalinement in waveguide joint.

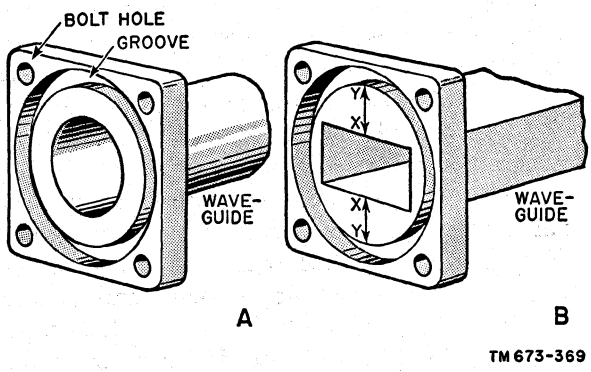


Figure 97. Choke fittings.

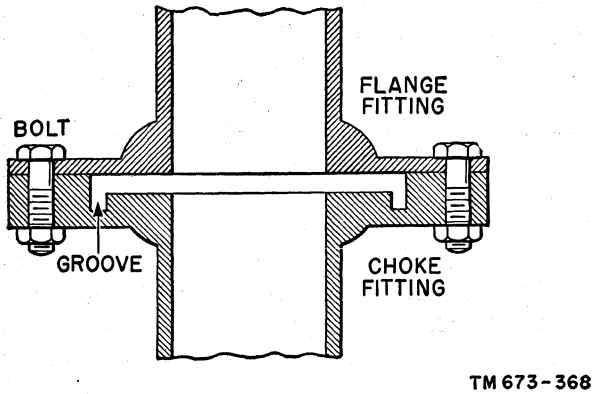


Figure 96. Choke joint.

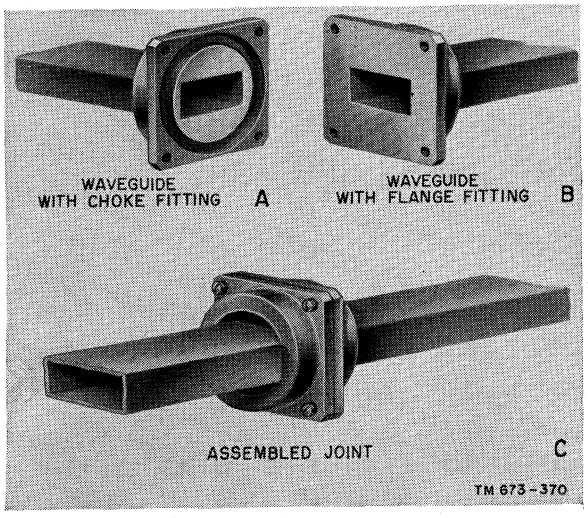


Figure 98. Rectangular waveguide joint.

the adjacent section with a flange fitting; and C shows the assembled joint with bolts in place.

c. Wobble Joint in Waveguide. It is frequently desirable to secure a waveguide transmission line to the structure of a ship or an aircraft while the unit which feeds the waveguide (a receiver, a transmitter, or a combination of both) is shock-mounted. The r-f unit then will be free to move slightly in relation to the waveguide. It is possible to make a connection to the waveguide by using a flexible section, but the usual method uses a *wobble joint.* A wobble joint is nothing more than a conventional choke joint in which the choke and flange fittings are separated by a distance of approximately $\frac{1}{16}$ inch. It is possible to move one section of the wobble joint in relation to the other for a considerable distance without causing any ill effects, because the presence of the choke groove creates a smooth-walled joint having electrical continuity. In a typical wobble joint, shifts of $\frac{1}{32}$ inch in any direction and rotations of 2° about any axis are permissible. It is customary to use a wobble joint with rectangular waveguide. If circular waveguide is used instead, the joint is electrically identical with the rotating joint described in paragraph 67 and either section may rotate through any angle about its electrical axis. Various positions of a wobble joint are shown, with some exaggeration, in figure 99.

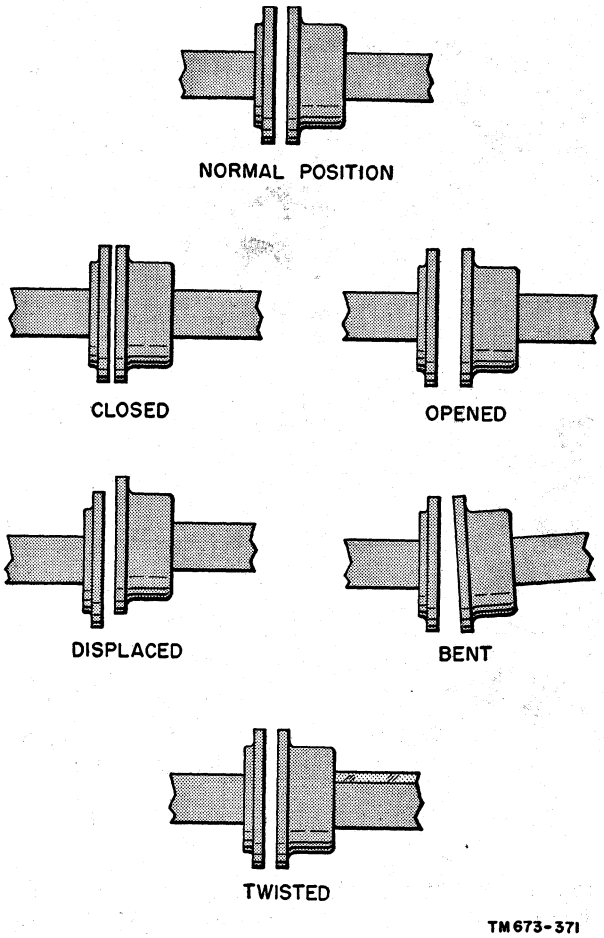


Figure 99. Various positions of wobble joint in rectangular waveguide.

69. Bends, Twists, and Flexible Sections

Special sections of coaxial line or waveguide are used when it is necessary to run a transmission system around corners. These are usually bends of various sorts and, in rectangular waveguide, they are twisted sections. A number of representative types are described below.

a. Coaxial Line Bends. A of figure 100, shows a *smooth* bend in coaxial line and B shows a *right-angle* bend. The quarter-wave stub supports, previously described, can be seen clearly. A smooth bend should be used in preference to a right-angle bend whenever there is any choice in the matter, since it is less likely to produce reflected signals.

b. Waveguide Bends. Bends in rectangular waveguide may be classified as *E-bends* (in which the waveguide is bent *up* or *down*, in the plane of the electric field) and *H-bends* (in which the guide is bent *sideways*, in the plane of the magnetic field). A of figure 101, shows a smooth E-bend in X-band rectangular waveguide; B is a smooth H-bend; C is a smooth bend in circular waveguide. The operation of bends of this type depends on the fact that any signal in the waveguide excites all the modes that are not suppressed. As long as the bend is smooth and gradual, very little reflection is produced. When it is necessary for the waveguide to bend sharply, a mitered bend is used. In figure 102, A is a mitered E-bend, and B is a mitered H-bend. Mitered bends rarely are used with circular waveguide. The operation of a mitered bend depends on reflection from the 45° wall, and this inevitably sends some of the signal back along the waveguide, as shown in figure 103. A smooth bend should be used instead of a mitered bend whenever there is a choice.

c. Twists. There is no occasion to use a twisted section of coaxial line because there is no special plane like the plane of polarization in a rectangular waveguide. For the same reason, there is no

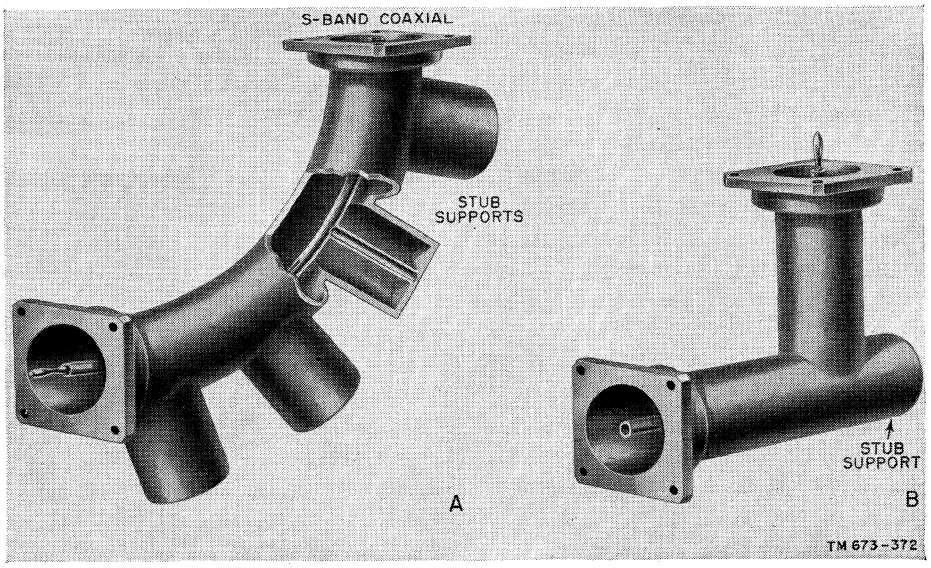


Figure 100. Coaxial line bends.

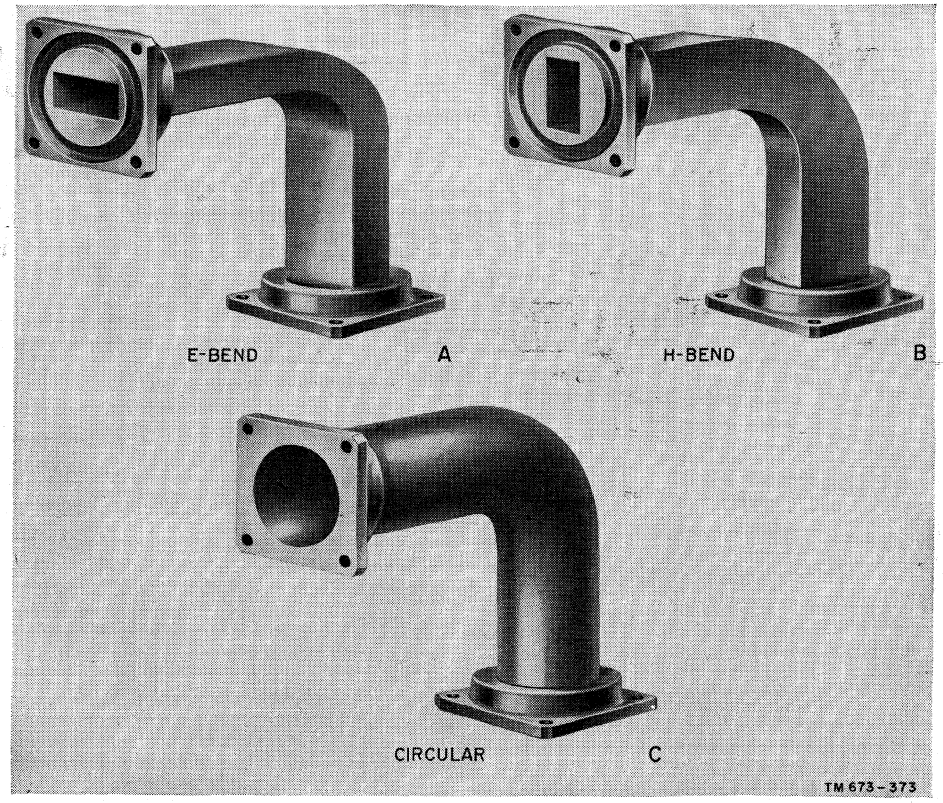


Figure 101. Smooth waveguide bends.

occasion to use a twisted section with circular waveguide operating in the $TM_{0,1}$ mode. It sometimes happens that it is necessary to change the plane of polarization of rectangular waveguide, and a twisted section is used for this purpose. Figure 104 shows a twisted section of X-band rectangular waveguide, and figure 105 shows an installation using twisted sections.

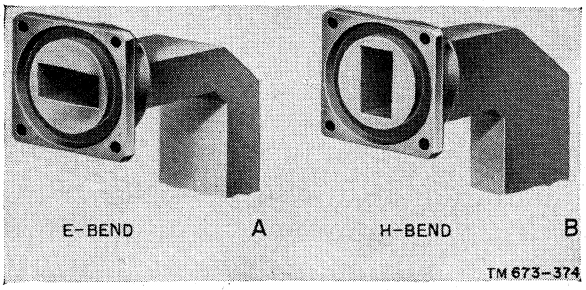


Figure 102. Mitered waveguide bends.

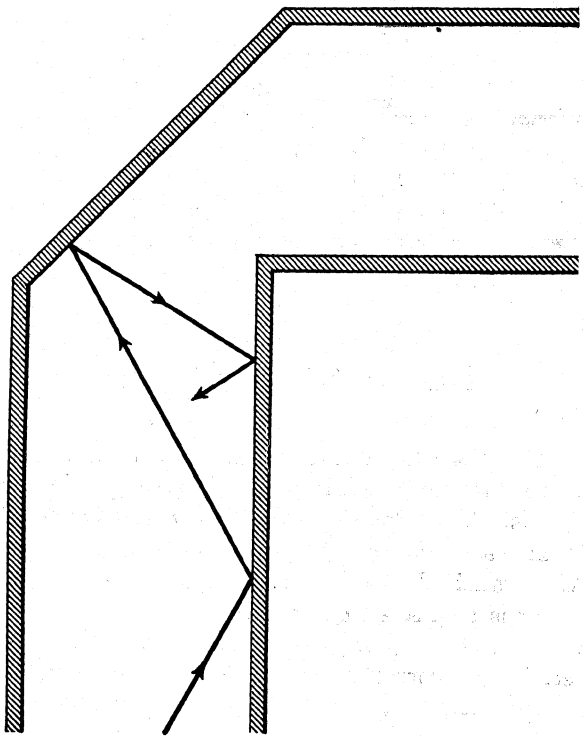


Figure 103. Reflected signal from mitered bend.

d. Flexible Sections. Flexible coaxial line is familiar to everyone with experience in the field of communications. Its use at microwave frequencies is avoided wherever possible because it has much higher losses than either waveguide or rigid coaxial line. Flexible waveguide is available, but its use is limited for the same reason—high losses. One type of flexible waveguide uses a woven conductor like shield braid. This is formed in a circular or rectangular tube and surrounded by a molded rubber tube of the same shape. The rubber sheath serves to hold the flexible conductor in the correct shape and maintains the desired circular or rectangular form of the waveguide. Another type consists of a spiral of flat metal ribbon, each turn being in electrical contact with its neighbors. As before, a rubber sheath surrounds the conductor and holds the turns in the proper relation to each other. A third type of flexible waveguide makes use of a number of choke fittings. These are held in a rubber sheath and form a very flexible guide that is free from troubles resulting from poor electrical contact. In general, flexible waveguide is avoided when a rotating joint or a wobble joint can be used instead. One type of flexible waveguide is illustrated in figure 106.

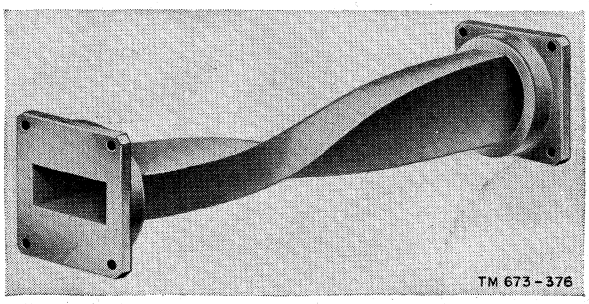


Figure 104. X-band waveguide twist.

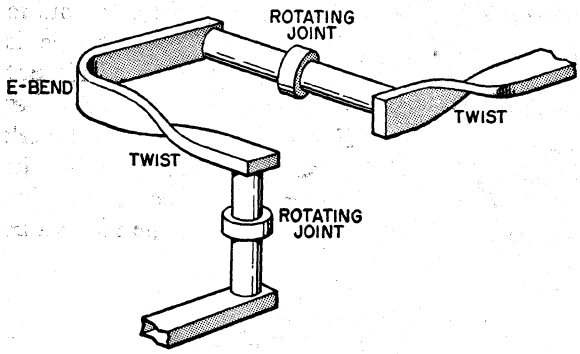


Figure 105. Waveguide installation with twists.

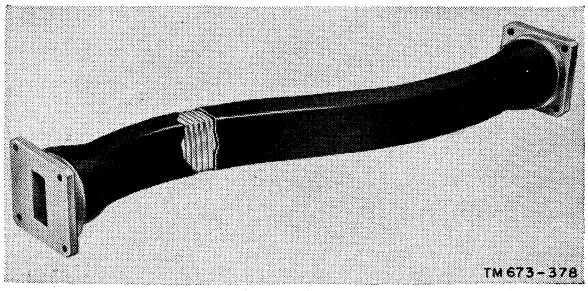


Figure 106. Flexible X-band waveguide.

70. Tee Junctions

Tee junctions in coaxial line have been described in connection with the stub tuner (par. 58*a* and fig. 60). The quarter-wave stub support is another example of a tee junction. Two types of tee junction are used with rectangular waveguide. Both of these, the E-junction and the H-junction, were described in paragraph 59*c* and shown in figure 73. Tee junctions seldom are used with circular waveguide. In addition to the junctions already described, there is an interesting double junction called a *magic tee* (fig. 107). This consists of a length of rectangular waveguide with an E-junction *and* an H-junction added at the same point. If a signal is fed in at A or at B, there will be some output from each of the other branches of the tee. If, however, a signal is fed in at C or at D there will be output *only* from branches A and B. A of figure 108 shows what happens when there is an input from branch C. The electric field is in opposite directions in the right and left portions of branch D and attempts to excite either a *TM* mode or a $TE_{2,n}$ mode. Both of these are suppressed by choice of the waveguide dimensions, and no energy is propagated down the branch. B illustrates what happens when there is an input at branch D. Here, the electric field attempts to excite the $TE_{0,2}$ mode in branch C, but this is suppressed as before. The junction will not behave in precisely the desired way *unless* branches A and B are flat. If any signal is reflected back in either of these branches, it will feed all of the other branches and there will be some coupling between branches C and D. A magic tee can be used in place of a directional coupler, and it is somewhat more sensitive in detecting reflected signals. The direct signal is fed *in* at C and *out* to the remainder of the transmission system from A. Branch B is terminated properly, and a detector is connected to branch D. So long as the transmission system is flat (and branch A is terminated correctly), there will be no signal at D. If, however, there is any reflected signal, a

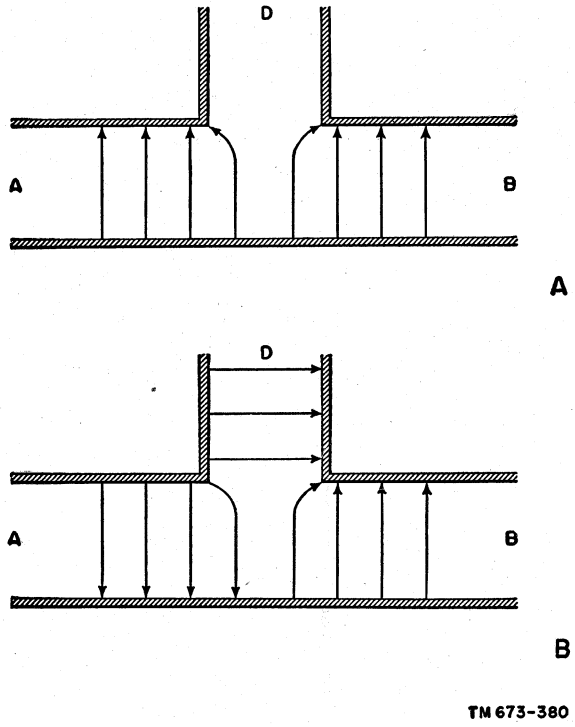


Figure 108. Operation of magic tee.

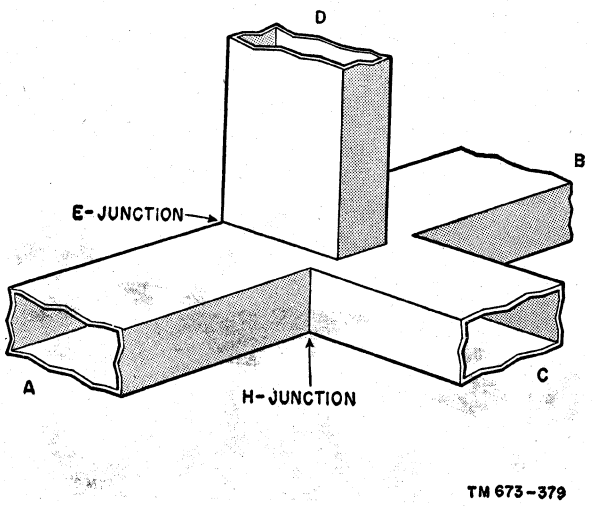


Figure 107. Magic tee.

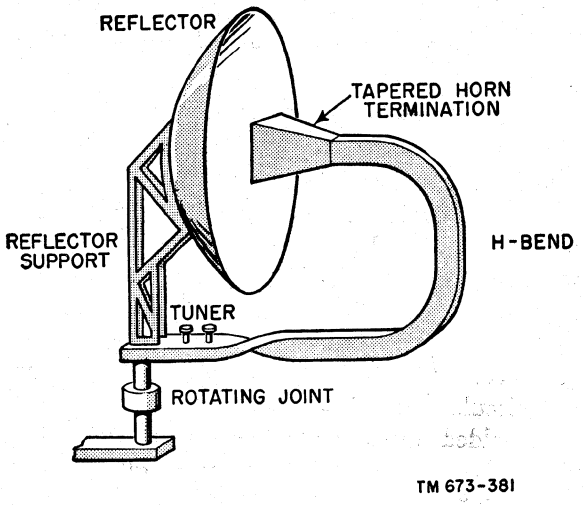


Figure 109. Tapered horn termination.

portion of it will reach the detector through branch D.

71. Coupling to Antenna

A transmission system, whether coaxial line or waveguide, usually runs from a transmitter to an antenna or from an antenna to a receiver. Any consideration of transmission systems must, therefore, include some discussion of the devices used for coupling between the antenna and the transmission line. This discussion will be deferred until chapter 7, which deals with microwave antennas. For the moment it is sufficient to note that there is always the problem of impedance matching at the antenna coupling. This problem frequently is met by the use of a preplumbed tapered section, and a tuner often is added for more precise matching. One example of a tapered matching unit is the horn termination shown in figure 109. Others are discussed in chapter 7.

Section V. OTHER WAVEGUIDES

72. Use of Solid Dielectric

Waveguides, instead of being *filled* with air, sometimes are filled with a solid dielectric. This increases the losses tremendously, but also raises the power-handling capacity by permitting the use of stronger electric fields because the breakdown voltage—the voltage at which an arc is formed—is higher for a good, solid dielectric than for air. This is the same effect that permits an oil-filled capacitor to operate at a higher voltage than an air capacitor. The behavior of a *filled waveguide* is the same as that of a conventional air-filled waveguide, except that, in all calculations, λ_a must be replaced by λ_d (the wavelength of the signals in the dielectric) and λ_d is equal to $\lambda_a \sqrt{k}$ where k is the dielectric constant of the material. Another application of solid dielectric is in the *dielectric waveguide*, which is nothing more than a solid rod of dielectric. This acts like a waveguide because signals traveling in the dielectric are reflected at the boundary. Reflection takes place because of the difference between the impedance of the dielectric and that of the surrounding space; the reflection at an impedance discontinuity is similar to the reflection at an impedance mismatch on a transmission line. The dielectric waveguide seldom is used. Not only does it have high losses; it also radiates to some extent because reflection at the boundary is not complete.

73. Single-Conductor Surface Wave Transmission Line

Paragraph 46 mentioned the use of a single conductor as a guiding surface for microwaves. A transmission line of this sort generally is called a *surface waveguide* or, more briefly, a G-line, from the initials of its originator, G. Goubau. It is described briefly in order to emphasize the fact that a single-conductor guiding structure need not be a hollow waveguide.

a. G-Line. The surface waveguide, or open waveguide that has been thoroughly investigated, is the conventional two-wire line. This type of line is useful at frequencies up to and including the vhf range. The small spacing required between the wires at higher frequencies and the difficulty of maintaining perfect symmetry to prevent excessive radiation make the two-wire line less useful at higher frequencies. The single-conductor line as an open waveguide is of particular interest, since it appears to be the simplest waveguide, and is most useful in the microwave ranges. Open waveguides have two advantages: They are less bulky and less expensive, and, since the field is not confined to a small area, as in a coaxial line or waveguide, the field density is low. This, in turn, has the effect that, under comparable conditions, the resistive and dielectric losses of open guides are much smaller than those of closed guides. The only practical wave mode that is guided by a single conductor is a radially symmetrical transverse magnetic mode. The radial extension of the field, by reducing the phase velocity, can be controlled by modifying the conductor surface. The simplest modification of the conductor surface is the application of a dielectric layer such as ordinary enamel or polyethylene. A very thin layer of dielectric will reduce the extension of the field considerably. Since the part of the energy that is propagated within the dielectric layer is small, the loss resulting from the dielectric material is almost negligible. A practical method of launching the surface wave on a single conductor becomes evident if the field of this wave (fig. 110) is compared with the field of a wave in a coaxial line. Starting with a coaxial-

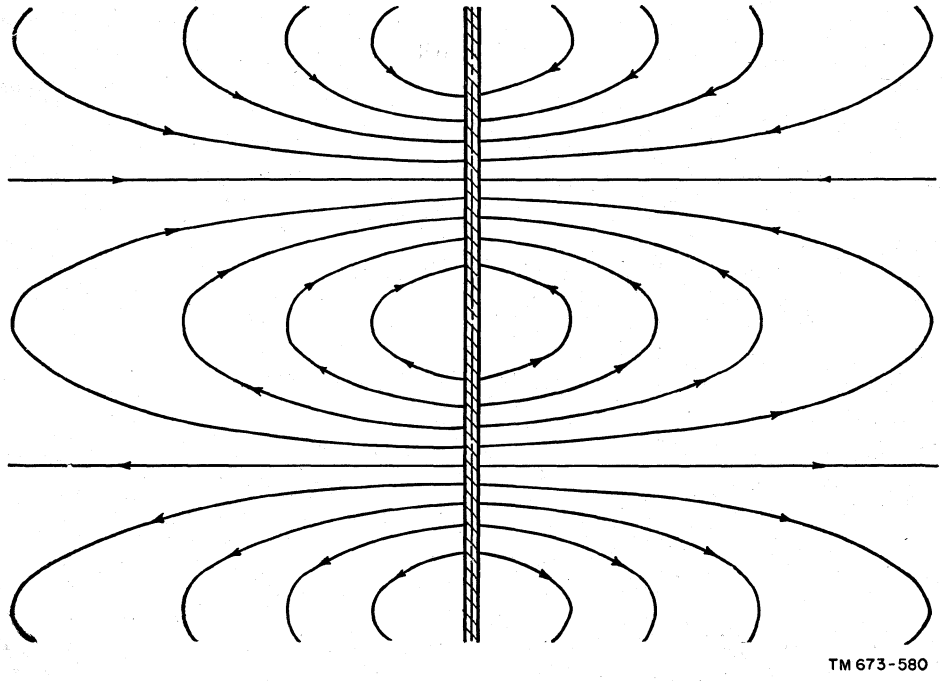
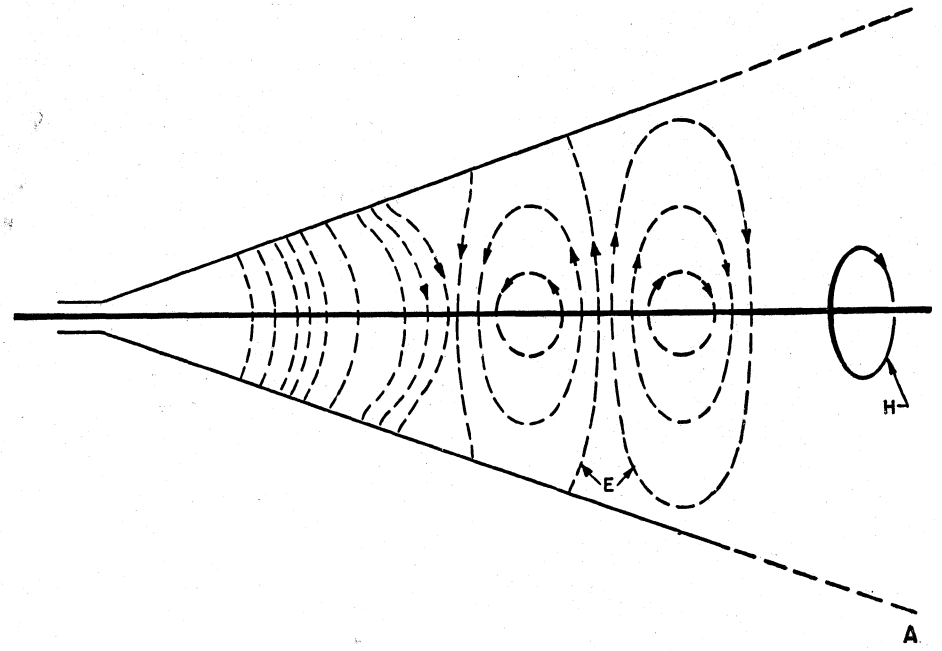
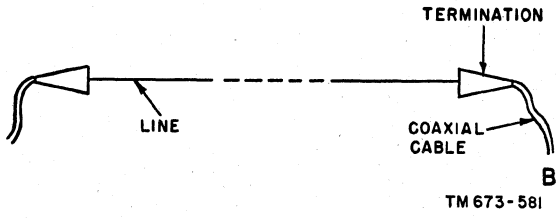


Figure 110. Single-conductor, open-guide field pattern.





A. Launching of surface wave.

B. Sketch of surface wave transmission line.

Figure 111. Surface wave.

line section, with the inner conductor having a dielectric coating, or layer, and the outer conductor of the coaxial section gradually increased until it is so large that it no longer affects the field very much, the field distribution becomes that of the open waveguide. The field lines in A of figure 111 give a rough idea of how the surface wave develops. The complete set-up for a surface wave transmission line is shown in B. The single-conductor, surface wave line is useful throughout the microwave frequency range. The line must be kept straight and clear of field obstructions as determined by the diameter of the launching horn. In addition, the line must be free of bends and kinks to prevent radiation loss.

b. Single-Wire Feed. Half-wave antennas at low frequencies sometimes are fed a short distance from the center by single wire. Of all the feeder systems available, this is probably the least desirable because the energy is transported by a radiating wave and its efficiency is low as a result of radiation. The radiating wave is like that on a long-wire antenna and not like the wave launched in the G-line at microwave frequencies, which is a nonradiating mode.

74. Coupling Between G-Line and Waveguide

A convenient method of feeding a G-line is from a section of waveguide, and the energy is received from the G-line with another section of waveguide. The coupling at each end of the G-line is the same, and consists of a tapered horn which matches a coaxial section to the surface waveguide. In order to support the surface guide, the coaxial section is made short and it is coupled to rectangular waveguide. The probe which couples the rectangular guide to the coaxial section then is used as a support for the G-line conductor (fig. 112).

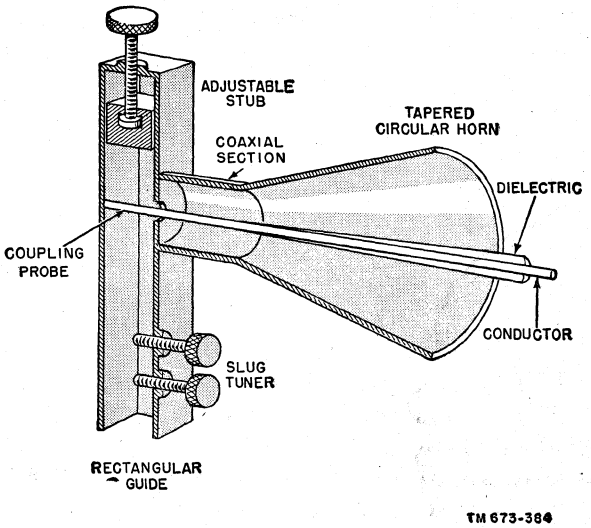


Figure 112. Coupling to G-line.

Section VI. PRACTICAL CONSIDERATIONS

75. Typical Transmission System Tuning

Most of the components used in microwave transmission systems are described in the earlier sections of this chapter. A portion of the transmission system of a microwave radar is shown in figure 113. This is a simple transmission system which shows how the various elements operate together. Although the system does not use all of the components that have been described, it is sufficiently complete to be a good example. The illustration shows a waveguide system, but the description of tuning procedure is directly applicable to a coaxial-line system, except that slotted sections are used instead of directional couplers. The remainder of the transmission system, which connects to the transmitter and receiver, uses devices called TR (transmit-receive) and ATR (anti-transmit-receive) tubes. These remove the receiver from the system while the transmitter operates, and also remove the transmitter from the system during the periods when the receiver is operating. The TR and ATR tubes, which act as switches, are mounted in resonant cavities. This portion of the system is discussed in chapter 4.

a. Tuning Antenna Feed. The entire tuning process is carried out with the transmitter operating and with the antenna pointed at the sky. Pointing the antenna away from any possible target prevents entrance into the system of all signals traveling *from* the antenna, except those that result from mistuning and consequent reflection. The antenna feed is tuned by the tuner immediately adjacent to the tapered horn. An indicator is connected to the directional coupler marked X (fig. 113) and the tuner is adjusted so that the output of the coupler drops to zero. The transmission system is then flat from the antenna end of the elevation joint to the antenna.

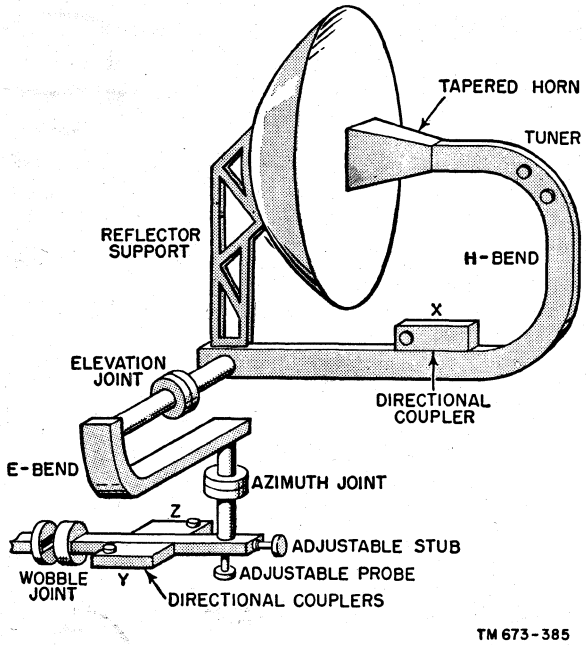


Figure 113. Transmission system with joints and adjustments.

b. Tuning Rotating Joints. *After* the antenna feed has been tuned, the indicating device is transferred to the directional coupler marked Y. The adjustable stub and adjustable probe are set in positions for which the directional coupler has no output. The system then is flat from the lower end of the azimuth joint back to the next discontinuity in the line. (This is probably the point where the line branches to connect to the receiver and the transmitter.) When this adjustment has been made, the system will be free from reflections caused by impedance mismatch, except in the portion between the antenna end of the elevation joint and the lower end of the azimuth joint. If the three nonadjustable couplings (between rectangular and circular guide) are preplumbed, there will be very little reflection. It would, of course, be possible to add more adjustments, but it probably would not be desirable, since only a little improvement could be obtained.

c. Further Adjustments. The indicating device may be transferred to the directional coupler marked Z in order to read the power input to this portion of the system.

76. Care of Transmission Systems

Almost all microwave transmission systems are filled with dry air under pressure and are sealed and, therefore, they require very little maintenance. From time to time they may be taken apart for inspection, and they must be treated with care to avoid bending or denting the walls. Any dirt, moisture, or corrosion within a waveguide or a coaxial line will increase the attenuation markedly. Any dents or scratches also will increase the losses and will increase the tendency toward arcing, thus limiting the power-handling capability of the transmission system. Since most waveguide is made of brass or aluminum, both of which are relatively soft, protection of the guide is a very real problem. Microwave test equipment frequently is plated with silver or gold to make it corrosion-resistant. It is, however, just as vulnerable to moisture and to physical damage as ordinary unplated waveguide. Waveguide is fabricated to extremely close tolerances, and finished inside to extreme smoothness. An attempt to repair the waveguide sections or components is, therefore, inadvisable if replacement parts can be obtained. A proper attitude toward a microwave transmission system is summed up in three sentences—

KEEP IT CLEAN!

KEEP IT DRY!

DON'T DENT IT!

77. Summary

a. Electromagnetic waves that travel through space, or through a uniform medium, are called unguided waves. Those that travel along a guiding structure, which defines the path followed by the waves, are called guided waves. At frequencies below the microwave region, the usual guiding structure is a conventional transmission line—coaxial, two-wire, or one-wire with ground return. At microwave frequencies, the usual guiding structure is a hollow waveguide, although rigid coaxial line is used at frequencies up to 3,000 mc per second, and surface waveguide sometimes is used at higher frequencies.

b. Hollow waveguide, used as a guiding structure or transmission line at microwave frequencies, consists of metal tubing. It is usually either rectangular or circular in shape.

c. The behavior of any transmission line may be considered on the basis of voltages and currents or from the standpoint of electric and magnetic fields. With conventional transmission lines, the voltage-current approach is much simpler than

the field approach, but with hollow waveguides it is easier to consider transmission of energy in terms of the fields involved. Even when the voltage-current approach is replaced by the field approach, such concepts as characteristic impedance, direct and reflected signals, and standing wave ratio are useful. Tuning a waveguide transmission system is, therefore, much the same as tuning a conventional transmission system.

d. Components used with coaxial line transmission systems at microwave frequencies are similar to those used at lower frequencies except that they are adapted for use with rigid coaxial line. Components used with hollow waveguide systems are essentially similar except for the absence of the inner conductor. Certain waveguide components, however, such as the magic tee and the directional coupler, have no low-frequency counterparts.

e. A hollow waveguide can carry energy in an infinite number of ways, each way having a different pattern of electric and magnetic fields. These ways are called modes, and every mode is either a *TE* (transverse electric) mode, a *TM* (transverse magnetic) mode, or a combination of the two. Associated with each mode is a cut-off wavelength and a corresponding cut-off frequency. A waveguide will not carry energy at a given mode unless the operating frequency is above the cut-off frequency for that mode. The cut-off frequency. and the wavelength depend on the dimensions of the waveguide.

f. The mode having the lowest cut-off frequency, and the longest cut-off wavelength, is known as the dominant mode. The dimensions of rectangular waveguide ordinarily are chosen so that operation at the frequency in use is impossible except at the dominant mode.

g. Each mode has associated with it a particular guide wavelength, a phase velocity, and a group velocity. These depend on the dimensions of the waveguide, on the operating frequency, and on the mode of transmission. The guide wavelength is always longer than the wavelength associated with the same frequency for transmission through space. The phase velocity, the velocity with which the field pattern moves through the waveguide, is greater than the velocity with which signals travel through space. The group velocity, the velocity with which energy is transported through the waveguide, is less than the velocity of the signals in space.

h. Waveguide transmission systems are tuned in the same way as conventional transmission systems—by adjusting impedance-matching devices in order to minimize reflected signals.

i. Waveguide and waveguide components are made of metal and, therefore, are strong. They must, however, be handled with great care, and it is necessary to avoid denting the walls or scratching the interior surface. In addition, the inside of the waveguide must be kept clean and dry at all times.

78. Review Questions

a. Distinguish between guided and unguided waves.

b. What types of transmission systems are used at microwave frequencies?

c. What is the maximum length of coaxial line used in S-band systems? In X- and K-band systems?

d. Why is the use of coaxial line impractical in the K- and X-bands except for coupling purposes?

e. Is the electric field in a coaxial line directed from the positive to the negative conductor, or vice versa? To what is the strength of the electric field in a coaxial line proportional?

f. Describe the direction of the magnetic field in a coaxial line.

g. What does the term *waveguide* describe? What two types are in common use?

h. Which dimension of a rectangular waveguide is referred to as the height—the shorter or the longer?

i. Is the electric field in a waveguide parallel to the longer dimension of the guide?

j. If b is the width of the waveguide, λ the wavelength of the signals, and B the angle that the path of the electromagnetic waves makes with the sides of the waveguide, give the equation that expresses the relationship between them.

k. Can the angle B of the equation in question (*j*) be any angle?

l. How may the patterns of the electric and magnetic fields in waveguide be determined?

m. Does the electric field vanish at the side walls of a waveguide? Is the transverse magnetic field present at the side walls?

n. Can a waveguide transmit energy if the wavelength in air, λ_a, is less than the cut-off wavelength, λ_c?

o. What is the velocity with which energy is transmitted along waveguide called?

p. What is the velocity with which the field pattern travels along waveguide called?

q. What are the full names for the *TE* and *TM* modes?

r. Why should the largest possible waveguide that will allow suppression of the undesired modes be used?

s. How may the operating wavelength of microwave equipment be estimated when rectangular waveguide is used?

t. With reference to circular waveguide, to what does the *m* in $TM_{m,n}$ refer?

u. What is the approximate diameter of circular waveguide used in S-band systems?

v. What does a double-stub section in coaxial line do?

w. In terms of wavelength, through what range may a tuning stub be adjusted?

x. For what is a line stretcher used in coaxial line?

y. How does a tuning slug in waveguide cancel the signal reflected from the load?

z. When are window tuners useful?

aa. What is the purpose of a slotted section in waveguide?

ab. What signal is measured by a directional coupler?

ac. Why is a directional coupler desirable in a pressurized waveguide system?

ad. What is the purpose of an attenuator in a waveguide?

ae. What is a *preplumbed* system?

af. What is the chief disadvantage of a filled waveguide?

ag. What three sentences of three words each sum up the proper attitude toward the care of microwave transmission systems?

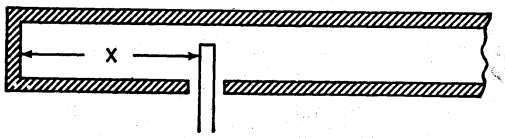
CHAPTER 4
CAVITY RESONATORS

Section I. THEORY

79. Introduction

A *cavity resonator*, sometimes called a *resonant cavity*, is nothing more than a space bounded by conducting walls—a hollow conductor. Cavity resonators are used, at microwave frequencies, in place of the L–C (inductance-capacitance) circuits and transmission line stubs employed at lower frequencies. The behavior of cavity resonators is discussed in detail in the following paragraphs. Before proceeding to this analysis, it is desirable to examine the behavior of a simple cavity, to gain an understanding of how it resonates.

a. Reinforcement of Signals. Consider the section of rectangular waveguide excited in the $TE_{0,1}$ mode by a small probe, which is shown in figure 114. The transmission system that supplies energy to the probe is not shown, since it does not enter the discussion. The waveguide extends a distance, X, to the left of the probe, where it is terminated by a conducting wall, and it extends indefinitely far to the right of the probe. The probe, of course, radiates signals to the right *and* to the left, and at any instant the signal moving to the right is made up of two parts. One part is simply the signal that the probe radiates to the right; the other part is the signal that has been reflected from the termination at the left—the signal that was radiated *to the left* by the probe a short time before. Now, if distance X is equal to $\lambda_g/4$, the signal radiated to the left must travel a half-guide-wavelength before it returns to the probe. Its phase is reversed on reflection at the conducting wall, so that it reaches the probe with a phase shift of 180°. Because it travels a half-guide-wavelength, its traveling time is exactly 1 half-cycle, and during this time the phase of the signal radiated by the probe also changes by 180°. If, therefore, distance X is one-quarter of λ_g (the guide wavelength), the signal radiated to the right, by the probe, will be reinforced by the reflected signal, originally radiated to the left by the probe, which is traveling to the right. If distance X is changed, the traveling time also is changed. The phase shift in the reflected signal, however, is always 180°, since it results from reflection at the terminating wall and *not* from traveling time. It follows, therefore, that the reflected signal will reinforce the signal radiated to the right whenever the traveling time is an odd number of half-cycles, since the phase shift at the probe during this time is, then, 180° plus some whole number of cycles. Reinforcement therefore occurs whenever $2X$ is an odd number of half-guide-wavelengths, or whenever X is an *odd* multiple of $\lambda_g/4$.

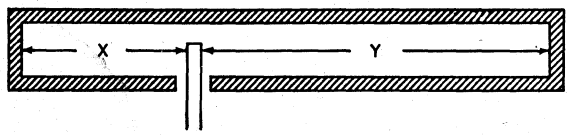


TM 673-434

Figure 114. Waveguide terminated by conducting wall

b. Simple Resonator. Now, consider the resonant cavity shown in figure 115. This consists of a section of waveguide similar to the one discussed above, except that it is terminated in both directions by conducting walls. If, now, *both* X and Y are odd multiples of $\lambda_g/4$, signals radiated to the right will be reinforced by those reflected from the left-hand wall, and signals radiated to the left will be reinforced by those reflected from the right-hand wall. The bulk of the signal, in each direction, consequently, is supplied by reflection from the opposite end, and only a small amount of energy needs to be supplied by the probe. For a given signal level in the cavity, the required input power is a minimum when X and Y are odd

multiples of a quarter-guide-wavelength. Alternatively, for a given signal input, the signal level in the cavity is greatest when X and Y are odd multiples of $\lambda_g/4$. The wavelengths or frequencies for which this is true are said to be characteristic frequencies of the cavity, and the cavity is said to resonate at these frequencies. The relation between X, Y, and λ_g is tabulated below for the four characteristic frequencies having the longest guide-wavelengths. Note that there are several possible positions for the probe at the higher characteristic frequencies. The longest guide-wavelength is the one for which the length of the cavity is $\lambda_g/2$, and the guide-wavelengths of the other characteristic frequencies are in the ratio 1, 1/2, 1/3, 1/4, and so on. Unfortunately, the relation between λ_g and λ_a is not the same for all of the characteristic



TM 673-430

Figure 115. Simple resonant cavity.

frequencies, and the wavelengths in *air* do not satisfy this simple relation, nor are the characteristic frequencies related by the reciprocals of these ratios (1, 2, 3, 4, and so on). Note that a second probe can be added to take energy *out* of the cavity. If, then, the input probe is excited at a constant signal level, the signal level in the cavity will be a maximum when the cavity resonates (the resonant frequency of the cavity may be changed by tuning it, or the frequency of the input signal may be changed to obtain resonance) and the energy transferred to the output probe also will be a maximum at resonance, since the output probe takes energy from the cavity.

X	Y	$X+Y$	λ_g
$\lambda_g/4$	$\lambda_g/4$	$\lambda_g/2$	$2\ (X+Y)$
$\lambda_g/4$	$3\lambda_g/4$	λ_g	$(X+Y)=2\ (X+Y)/2$
$3\lambda_g/4$	$\lambda_g/4$	λ_g	$(X+Y)=2\ (X+Y)/2$
$\lambda_g/4$	$5\lambda_g/4$	$3\lambda_g/2$	$2(X+Y)/3$
$3\lambda_g/4$	$3\lambda_g/4$	$3\lambda_g/2$	$2(X+Y)/3$
$5\lambda_g/4$	$\lambda_g/4$	$3\lambda_g/2$	$2(X+Y)/3$
$\lambda_g/4$	$7\lambda_g/4$	$4\lambda_g/2$	$2(X+Y)/4$
$3\lambda_g/4$	$5\lambda_g/4$	$4\lambda_g/2$	$2(X+Y)/4$
$5\lambda_g/4$	$3\lambda_g/4$	$4\lambda_g/2$	$2(X+Y)/4$
$7\lambda_g/4$	$\lambda_g/4$	$4\lambda_g/2$	$2(X+Y)/4$

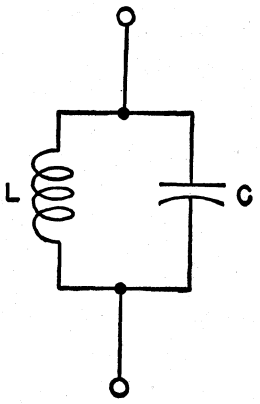
80. Principles of Resonant Circuits

Since the behavior of a cavity resonator resembles that of an L-C circuit as well as that of a transmission line section, it is convenient to introduce the detailed analysis of resonator behavior by examining the behavior of the two conventional circuits.

a. Resonant L-C Circuit. Consider the shunt-resonant, zero-resistance L-C circuit shown in figure 116. This consists of an inductor, L, in parallel with a capacitor, C. When an alternating voltage is applied to the terminals of this circuit, the phase of the current through the inductor is 90° behind the phase of the applied voltage. The phase of the current through the capacitor is 90° ahead of the phase of the applied voltage, and the two currents (through the inductor and the capacitor) therefore differ in phase by 180°. At any frequency, the two currents cancel each other to some extent; however, at the resonant frequency, f_r, the two currents are equal in magnitude (and opposite in phase). Consequently, they cancel each other completely and no current flows through the L-C circuit from the external source, although heavy circulating current flows *in* the circuit loop. The resonant frequency is defined by the relation,

$$fc=\frac{1}{2\pi\sqrt{LC}}$$

At its resonant frequency, the impedance of this circuit is infinite: It passes no current. In practice, the inductor has some resistance and the capacitor may have both lead resistance and leakage. The impedance at resonance in the practical case is, therefore, not infinite, but only very great.



TM 673-433

Figure 116. Resonant L-C circuit.

For the purpose of this discussion, it is sufficient to note that the impedance is *infinite*, or very high, at the resonant frequency.

b. Resonant Transmission-Line Section.

(1) The circuit shown in A of figure 117, is nothing more than a section of transmission line, of length X, terminated by a short circuit. The previous discussion of stub tuning makes it clear that there will be a reflected signal whenever an alternating voltage is applied to the input terminals of the shorted section of transmission line. The amplitude of the reflected signal will be equal to that of the direct signal and, at the termination, their phases will differ by 180°. This condition is imposed by the fact that the terminating impedance is zero. The voltage at the termination is zero, since the direct and reflected signals have equal amplitudes and differ in phase by 180°. The current, however, is not zero, since the currents of the direct and the reflected signals add directly; forward current for the direct signal flows in the same direction as backward current for the reflected signal. At a point a quarter-wavelength from the termination, the phase of the direct signal is advanced by 90°, and that of the reflected signal is retarded by 90°. The two signals, which differ in phase by 180° at the termination, are therefore in phase with each other at a point a quarter-wavelength from the termination. The voltages of the two signals then add, whereas the currents cancel each other (the forward currents of the direct and reflected signals flow in opposite directions). Since the current is zero, the impedance at a point a quarter-wavelength from the termination is infinite.

(2) The same condition obtains at all points an integral number of half-wavelengths from this point (points an odd number of quarter-wavelengths from the termination), since moving a half-wavelength along the line changes the phase of each signal by 180° and, therefore, leaves them in phase with each other. In practice, of course, the impedance is very high but not infinite. The behavior of this circuit, then, is described by the statement that the input impedance is *infinite* whenever the frequency of the applied signal is such that X is an odd number of quarter-wavelengths. Now, consider the circuit shown in B of figure 117, which is a section of transmission line terminated at each end by a short circuit, with input terminals connected at a suitable point. This will have an *infinite* input impedance if X is an odd number of quarter-wavelengths *and* Y is also an odd number of quarter-wavelengths. The total length of the section of transmission line, for an infinite input impedance at a suitable point, is, then, $(X+Y)$, which is an *even* number of quarter-wavelengths (since the sum of two odd numbers is an even number) or some multiple of a half-wavelength. Consequently, there will be a point (or a number of points) an odd number of quarter-wavelengths from each end of the section at which the impedance is *infinite*; this is true whenever the frequency is such that the total length of the section is an exact multiple of a half-wavelength. The characteristic wavelengths for which the input impedance is *infinite* are given by the relation

$$\lambda_k = 2(X+Y)/k,$$

where k is a positive integer (1, 2, 3, and so on), and where the wavelength is, as noted, the wavelength on the transmission line. The *dominant mode* of resonance is the one having the greatest wavelength. The characteristic wavelength for the dominant mode is, therefore

$$\lambda_1 = 2(X+Y)/1 = 2(X+Y)$$

The wavelength for the k mode, is, of course λ_k, and this is $1/k$ times λ_1. It follows, therefore, that f_k, the characteristic frequency of this mode, is approximately k times f_1. The characteristic frequencies are, then, very nearly exact multiples of characteristic frequency f_1 of the dominant mode. The harmonic relation (exact multiples) is precise if the ratio between wavelength on the transmission line and wavelength in air is the

same at all frequencies. The behavior of the circuit shown in B of figure 117, is described by the statement that its input impedance is *infinite* at every characteristic frequency. Each characteristic frequency, f_k, is associated with a characteristic wavelength, λ_k, which is given by the relation

$$\lambda_k = 2Z/k,$$

where Z is the length of the transmission line section.

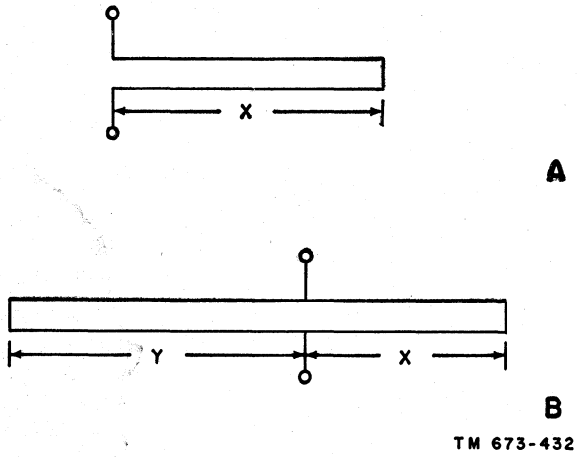


Figure 117. Resonant transmission line sections.

c. Resonant Cavity. It is impossible to write a simple expression for the characteristic frequencies of a *general* cavity resonator. If, however, the resonator consists of a section of waveguide terminated at each end by a conducting wall, it is precisely equivalent to the section of transmission line described in *b* above. Here, the characteristic guide wavelengths for which the input impedance of the resonator (at a suitable point) is infinite are given by the relation

$$\lambda_{gk} = 2Z/k$$

which may be translated as follows: the wavelength in the guide, λ_{gk}, is equal to twice the length of the waveguide, $2Z$, divided by k, which is any positive integer. Trouble arises as soon as and attempt is made to compute the wavelength in air and the corresponding frequency which are associated with a given λ_{gk}. In order to determine λ_{ak} from λ_{gk}, it is necessary to know the dimensions of the waveguide (its height and width if it is rectangular, its diameter if it is circular) in addition to the length of the section. It is also necessary to know the mode in which the waveguide is operating. Finally, the relation between λ_a and λ_g may be the same for more than one mode and, therefore, there may be two distinct field patterns for a given characteristic frequency. The whole problem is discussed in detail for a rectangular resonator in paragraph 81. For the moment, it is necessary only to point out that there is neither a *single* characteristic frequency, as in the L-C circuit, nor a *simple* series of characteristic frequencies, as in the transmission line section. It is stated frequently that a cylindrical resonator may be thought of as an infinite number of quarter-wave shorted stubs, as shown in figure 118. This implies that the characteristic frequencies are related to each other in the same way as those of the transmission line section described above. This explanation of resonator behavior completely neglects the existence of various modes in waveguides and is so oversimplified that it is practically of no value. The only constructive way of thinking about a cavity resonator is to consider it as a section of waveguide.

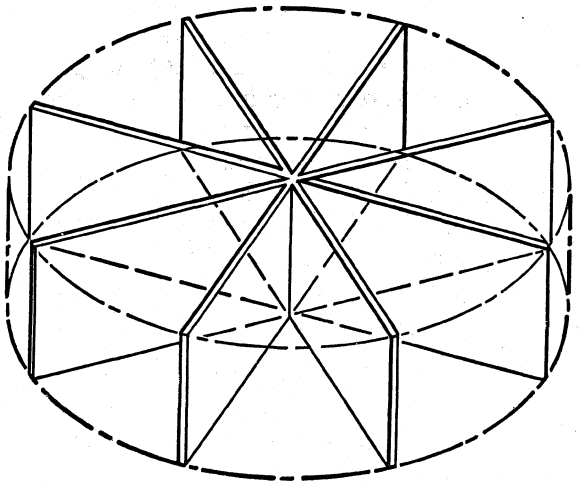


Figure 118. Characteristics of cylindrical resonator.

81. Rectangular Resonators

Since the mathematics of rectangular cavity resonators is based on the mathematics of rectangular waveguide, it is not difficult to understand. It is possible to discuss rectangular resonators in considerable detail, using simple mathematics, whereas this is not true of other resonators. It must be kept in mind that the cavity consists of a section of rectangular waveguide terminated at

each end by a conducting wall, which is assumed to be a perfect conductor. The problems of feeding energy to the waveguide section—the resonator—and taking energy out are discussed in paragraph 85.

a. Basic Theory.

(1) Consider a rectangular cavity resonator having a smallest dimension (height), a, an intermediate dimension (width), b, and a longest dimension (length), c. The use of c for the length of a cavity resonator should not be confused with its use for the velocity of a light wave or a radio signal. For the moment, consider this resonator as a waveguide section of length c, where the waveguide of which it is composed has a height a and a width b. Suppose, now, that the resonator is excited at a frequency for which

$$\lambda_g = 2c/p$$

where p is a positive integer. Because of the boundary conditions at a conducting surface, which require that the electric field *parallel* to the surface and the magnetic field *perpendicular* to the surface vanish *at* the surface, the signal in the waveguide will be reflected from the end of the cavity with a reversal of phase. The direct and incident field patterns, then, will be 180° out of phase and will, of course, be traveling in opposite directions.

(2) At the end of the cavity, the transverse electric fields of the direct and reflected field patterns in the waveguide will cancel each other, as will the longitudinal magnetic fields. The longitudinal electric fields and the transverse magnetic fields, however, will add, like the currents in the transmission line described in paragraph 80 because the forward direction for the direct signal is opposite from the forward direction for the reflected signal. At a quarter-wavelength from the end of the cavity, the direct and reflected field patterns are in phase. At this point, the transverse electric fields and the longitudinal magnetic fields add, and the longitudinal electric fields and the transverse magnetic fields cancel. At a quarter-wavelength farther from the end of the cavity, the earlier situation (the phase relation that is present at the end wall) is repeated. The result is a field pattern, formed by the direct and reflected signals, which is repeated every half-wavelength along the cavity. All of the distances above are measured, of course, in terms of the guide wavelength. If, now, the waveguide is operating in either the $TE_{m,n}$ or the $TM_{m,n}$ mode, and the length of the cavity is p times $\lambda_g/2$, the resulting field pattern in the cavity will be repeated m times along the a dimension, n times along the b dimension, and p times along the c dimension.

b. Various Modes of Operation. In this discussion, it has been assumed that the rectangular cavity was operated by making waveguide signals travel from one end of the cavity to the other. It is, of course, possible to send signals *across* the cavity, or *between the top and the bottom.* The length of the waveguide section and the dimensions of the waveguide are tabulated below for a cavity of height a, width b, and length c.

Direction of waveguide signals	Length of waveguide section	Height of waveguide	Width of waveguide
End to end	c	a	b
Side to side	b	a	c
Top to bottom	a	b	c

c. Operation With TE and TM Signals. Suppose that a rectangular cavity, of height a, width b, and length c, is excited by sending $TE_{m,n}$ signals *along* the cavity, from end to end, at a wavelength for which λ_g is equal to $2c/p$. At this point, consider what happens if the $TE_{m,n}$ signals are replaced by $TM_{m,n}$ signals having the same guide wavelength, so that λ_g remains equal to $2c/p$. It is assumed that m and n are the same for the TE and the TM signals. $TE_{3,5}$ signals, for example, are replaced by $TM_{3,5}$ signals. As explained in chapter 3, a waveguide pattern having a guide wavelength

$$\lambda_g = \lambda_a/\cos A \cos B$$

can be built up from four sets of vertically polarized waves the paths of which make angles A with the top and bottom of the waveguide and angles B with the side walls of the guide; also, a similar pattern can be built up from horizontally

polarized waves having the same λ_a. If the same values of m and n are used (par. 55) both patterns will have the same guide wavelength. Although the angles A and B are different for the two patterns, the product $\cos A \cos B$ is the same for both. Thus, if m, n, and λ_a are the same for both patterns, λ_g also will be the same. It has been pointed out that these patterns can be combined with one phase relation to obtain the pattern of a $TE_{m,n}$ mode and with another phase relation to obtain the pattern of a $TM_{m,n}$ mode. It follows, therefore, that a $TE_{m,n}$ mode and a $TM_{m,n}$ mode have the same guide wavelength if m, n, and λ_a are the same for both modes. Alternatively, if m, n, and λ_g are the same for both modes, they will have the same λ_a. It is clear, then, that λ_a is unchanged when the $TE_{m,n}$ signals, having λ_g equal to $2c/p$, are replaced by $TM_{m,n}$ signals having the same m, n, and λ_g. It is shown in appendix III that λ_a for either case is given by the relation,

$$\lambda_a = 1/\sqrt{(m/2a)^2+(n/2b)^2+(p/2c)^2}.$$

d. Characteristic Wavelengths. The characteristic wavelength (the value of λ_a) was calculated above for a cavity of height a, width b, and length c excited by $TE_{m,n}$ or $TM_{m,n}$ signals traveling *along* the cavity, from end to end, and having a wavelength for which λ_g is equal to $2c/p$. This mode of exciting the cavity produces a field pattern that is repeated m times along the a dimension, n times along the b dimension, and p times along the c dimension. An examination of the expression for λ_a shows that it involves a, b, c, m, n, and p, but makes no mention of the direction in which the waveguide signal travels. It follows, therefore, that λ_a will have the same value (in a given cavity) for *every* mode of operation in which the field pattern is repeated m, n, and p times along the a, b, and c dimensions, respectively. The characteristic wavelengths are, then, those for which

$$\lambda_a = \lambda_{m,n,p} = 1/\sqrt{(m/2a)^2+(n/2b)^2+(p/2c)^2}$$

where there may be more than one mode of operation of the cavity at a given characteristic wavelength. The longest characteristic wavelength is, to some extent, equivalent to the cut-off wavelength of a waveguide, since the cavity cannot resonate at longer wavelengths. The parallel is not exact, however, because the waveguide operates at any frequency above cut-off, whereas the cavity resonates only at its characteristic frequencies.

e. Terminology of Modes.

(1) Modes in a rectangular cavity resonator are described as TE modes if they are excited by TE waveguide signals and as TM modes if they are excited by TM waveguide signals. Each mode is described further by three subscripts which show how many times the field pattern is repeated along the height, width, and length of the cavity. There are, therefore, $TE_{m,n,p}$ modes and $TM_{m,n,p}$ modes. If the electric field is perpendicular to, for example, the top and bottom of the cavity, it is possible to have a pattern that is uniform along one dimension (in this case, the height) so that one subscript may be zero. Here, m is zero. A field that is perpendicular to two walls is, however, parallel to the other four. A pattern that is uniform one dimension, therefore, must vary along *both* of the other dimensions in order that the electric field may satisfy the boundary conditions by vanishing at the surface of each of the four walls to which it is parallel. It is impossible, on this account, to have a mode in which the field pattern is uniform along more than one direction, and there are no modes for which more than one subscript is zero.

(2) Field patterns that are uniform along one dimension are excited by waveguide signals like those of the $TE_{0,n}$ modes or the $TE_{m,0}$ modes in a rectangular waveguide and are, accordingly, TE modes. The cavity resonator, therefore, may operate in either a $TE_{m,n,p}$ mode or a $TM_{m,n,p}$ mode at a characteristic wavelength, $\lambda_{m,n,p}$, for which all of the subscripts are nonzero, and it may operate only in a $TE_{m,n,p}$ mode at a characteristic wavelength, $\lambda_{m,n,p}$, for which one of the subscripts is zero. There are no characteristic wavelengths for which more than one subscript is zero.

f. Field Patterns of Modes.

(1) The dominant characteristic wavelength—the longest one—is $\lambda_{0,1,1}$. The corresponding mode is the $TE_{0,1,1}$ mode, which has the field pattern shown in A of figure 119. This mode may be excited by a

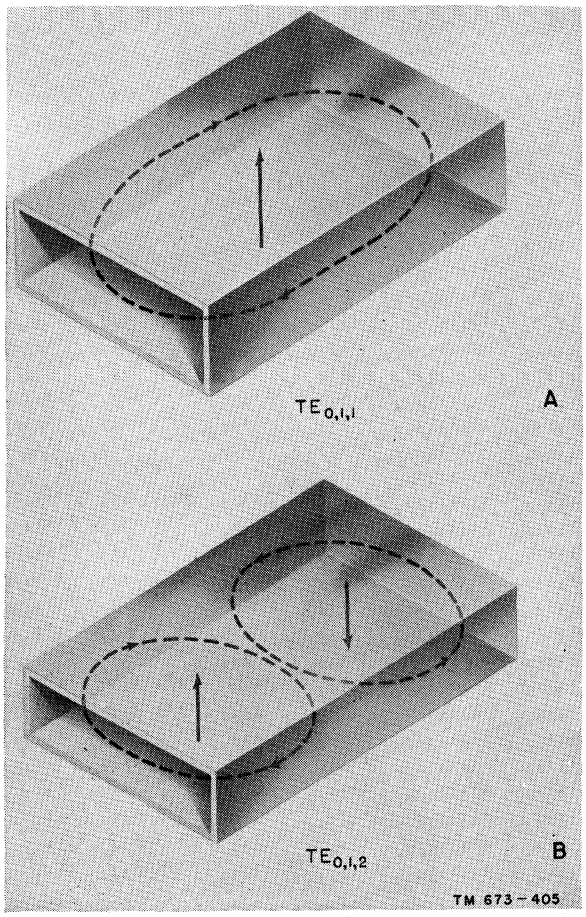


Figure 119. Simple resonator modes.

$TE_{0,1}$ signal traveling along the resonator, from end to end, or by a $TE_{0,1}$ signal traveling across the resonator from side to side. The field pattern is the same in both instances so that there is, in fact, only a single $TE_{0,1,1}$ mode. Similar patterns are obtained in the $TE_{1,0,1}$ mode, at $\lambda_{0,1,1}$, where the electric field is perpendicular to the *sides* of the cavity (in the $TE_{0,1,1}$ mode the electric field is perpendicular to the top and bottom of the cavity), and in the $TE_{1,1,0}$ mode, at $\lambda_{1,1,0}$, where the electric field is perpendicular to the *ends* of the cavity. All *TE* modes having one subscript equal to zero have field patterns that are repetitions of the simple $TE_{0,1,1}$ pattern. A single example, the pattern of the $TE_{0,1,2}$ mode, is shown in B.

(2) The $TE_{0,1,2}$ mode may be excited by a $TE_{0,1}$ signal traveling along the resonator or by a $TE_{0,2}$ signal traveling across the resonator. The pattern, however, is the same in each case. For any characteristic wavelength having one subscript equal to zero, there is a *single TE* mode at which the cavity may operate. There are several possible modes, however, for a characteristic wavelength having three nonzero subscripts. At $\lambda_{1,1,1}$, for example, the cavity may be exicted by a $TE_{1,1}$ signal traveling along the cavity from end to end. This develops a $TE_{1,1,1}$ mode, with the pattern shown in A of figure 120. Note that the electric field is parallel to the ends of the cavity. A second $TE_{1,1,1}$ mode may be developed by a $TE_{1,1}$ signal traveling across the cavity. This will have a similar pattern, but the electric field will be parallel to the sides of the cavity. Finally, a third $TE_{1,1,1}$ mode may be developed by a $TE_{1,1}$ signal traveling from the top of the

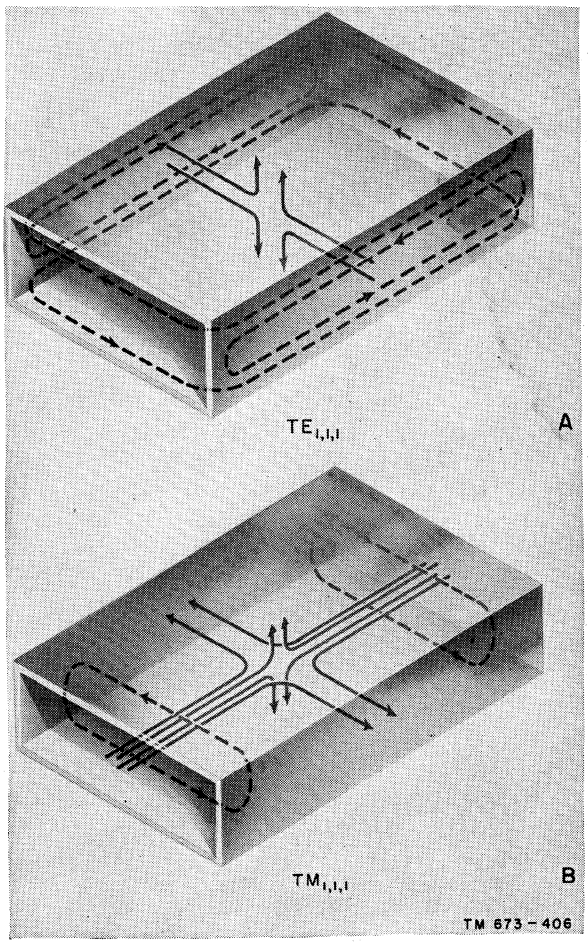


Figure 120. Higher resonator modes.

cavity to the bottom. Here, the electric field will be parallel to the top and bottom of the cavity. There are also, for the same characteristic wavelength, three $TM_{1,1,1}$ modes, all excited by $TM_{1,1}$ signals. In one, shown in B, the signals travel along the cavity, from end to end, and the magnetic field is parallel to the ends of the cavity. In the two other modes, the signal travels across the cavity, with the magnetic field parallel to the sides of the cavity, or it travels from top to bottom, with the magnetic field parallel to the top and botton of the cavity. The various patterns for characteristic wavelengths with subscripts greater than 1 are repetitions of the simple patterns. For any characteristic wavelength having three nonzero subscripts, there are three *TE* modes and three *TM* modes in which the cavity may operate.

82. Cylindrical Resonators

A detailed discussion of cylindrical resonators, like a discussion of circular waveguide, involves extremely complicated mathematics beyond the scope of this manual. Therefore, the subject is discussed only briefly. As with a rectangular resonator, the characteristic wavelengths of a cylindrical resonator are distinguished from each other by three subscripts, m, n, and p. There is, unfortunately, no simple expression that relates $\lambda_{m,n,p}$ to the dimensions of the resonator. The subscripts, m and n, like those used to describe a mode in circular waveguide, refer to the number of times that the field pattern is repeated around the circumference of the cylinder, m, and the number of times that the pattern is repeated from the axis of the cylinder to its circumference, n. The third subscript, p, refers, as in the rectangular resonator, to the number of times that the field pattern is repeated along the length of the resonator, along the axis of the cylinder. The simplest mode in a cylindrical resonator is the $TE_{1,1,1}$ modes, shown in A of figure 121, which corresponds to the $TE_{0,1,1}$ mode in a rectangular resonator. Another simple mode is the $TM_{0,1,0}$ mode, in B, which corresponds to the $TE_{1,1,0}$ mode in a rectangular resonator, and is seen frequently in resonators that have a modified cylindrical form. Among such resonators are the cavity of an X-band

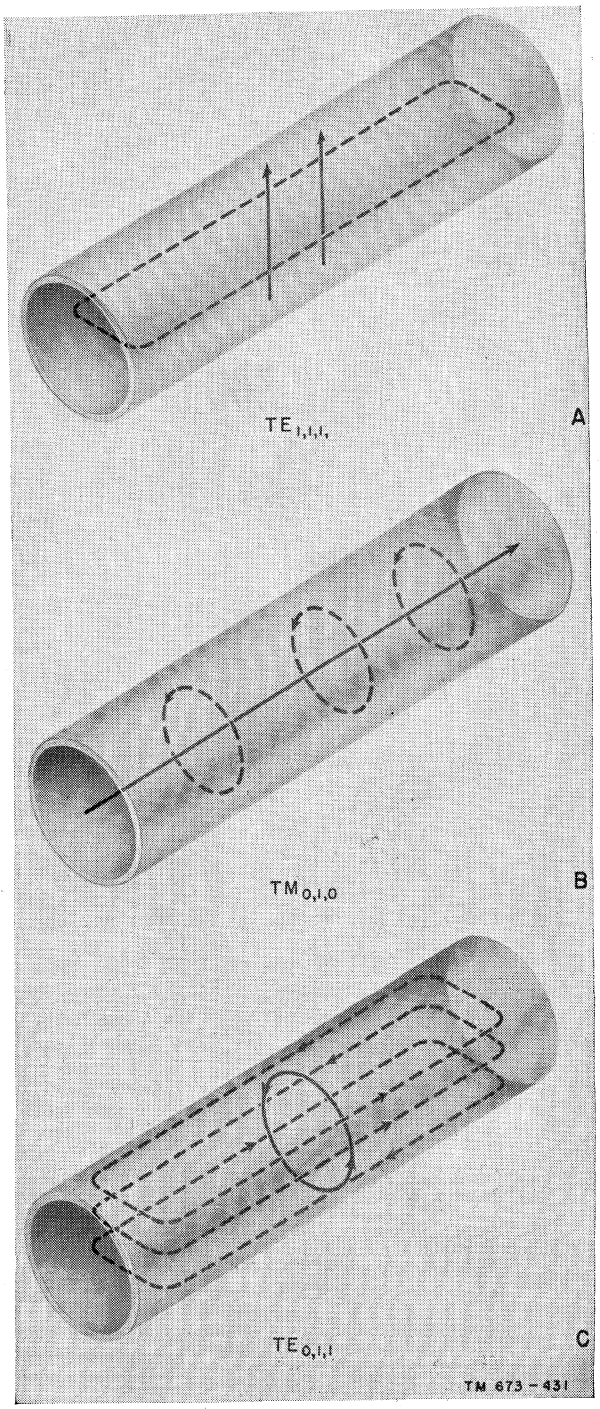


Figure 121. Field patterns in cylindrical resonators.

klystron, mentioned in the following paragraph and described in detail in chapter 5, and the cavities used for TR and ATR switches (par. 88) Another important mode is the $TE_{0,1,1}$ mode shown in C. A cavity resonator operated in this mode has extremely low losses. Its input impedance at resonance is, therefore, extremely high and

the variation in impedance with change in frequency is very sharp. Microwave wavemeters are often cylindrical cavities operating in this mode.

83. Other Resonators

Cavity resonators are seldom simply rectangular or cylindrical. The rectangular resonator has been discussed in detail to illustrate the basic theory behind all cavity resonators, and the cylindrical resonator has been described briefly because most actual resonant cavities operate in modes similar to the $TM_{0,1,0}$ and $TE_{0,1,1}$ modes of a cylindrical resonator. Three representative *re-entrant* cavities are illustrated in figure 122. The term *re-entrant* indicates that a portion of the wall protrudes into (or enters) the space within the cavity. A shows the resonant cavity of an X-band reflex klystron, B shows a TR cavity, and C represents a re-entrant wavemeter cavity. All three operate in modes similar to the $TM_{0,1,0}$ mode of a cylindrical resonator. In each, the direction of the electric field is indicated.

84. Cavity Resonator Tuning

One method of tuning a cavity resonator is to change its dimensions by moving one of its walls in much the same way that the termination of a tuning stub is moved (par. 58*c*). This actually is done in some wavemeter cavities. Another method is to distort the cavity by pressure at appropriate points. A third method is to distort the electric or the magnetic field, thus changing the effective dimensions of the cavity. The wavemeter cavity shown in C of figure 122, is tuned in this manner: Changing the position of the pin alters the electric field. The magnetic field in a cavity usually is changed by inserting plugs which extend a short distance from the wall. For large changes in the magnetic field, tuning paddles frequently are used, instead of plugs. A tuning plug and a tuning paddle are shown in the wall of a cavity in A of figure 123. B shows

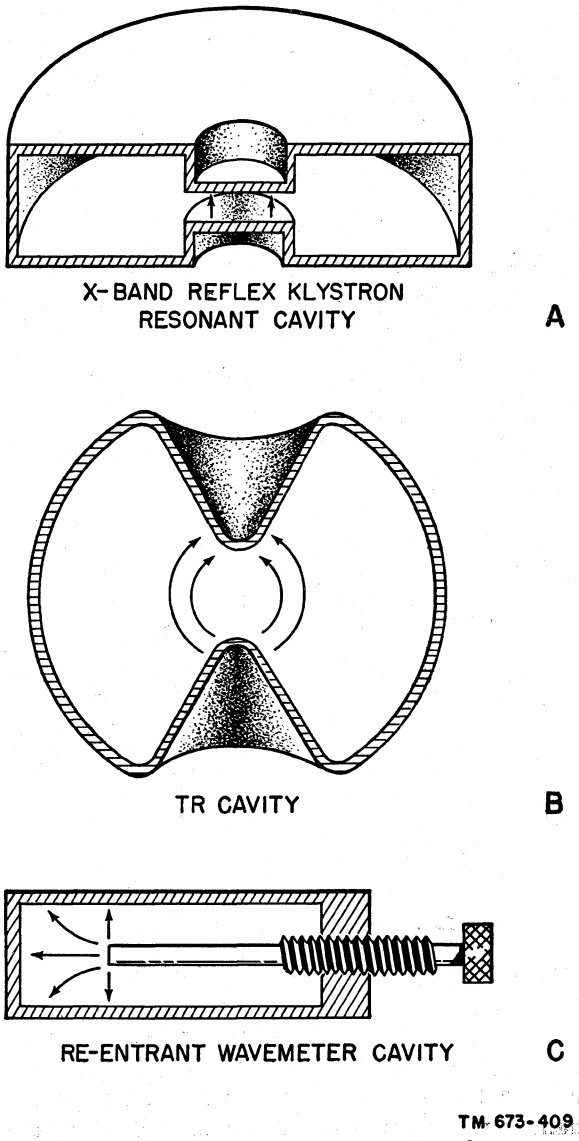


Figure 122. Cavity resonators of various shapes.

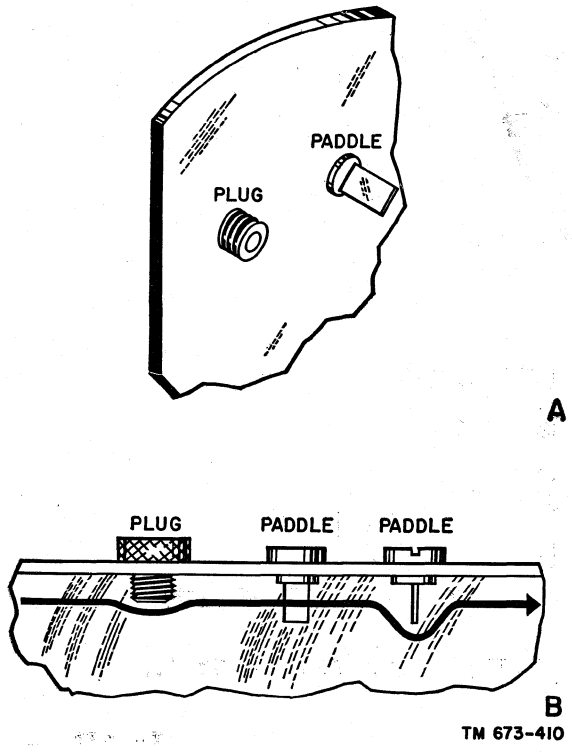


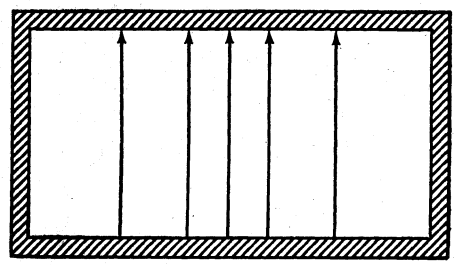
Figure 123. Tuning plug and paddle.

the consequent distortion of the magnetic field. The operation of the plug is obvious. The operation of the paddle also is simple: So long as the plane of the paddle is parallel to the direction of the magnetic field, it has no effect; when the plane of the paddle is perpendicular to the direction of the field, the distortion is greatest.

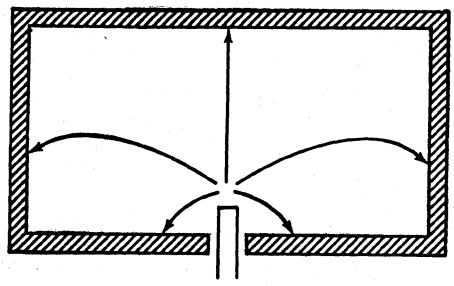
85. Coupling to Cavity Resonators

Energy may be fed into or taken from a cavity resonator in three ways—by using a probe, a loop, or a slit.

a. Probe Coupling. Probe coupling seldom is used with resonant cavities. If the probe extends deeply into the cavity, like the probe used to feed a waveguide, it is likely to load the cavity so heavily that the resonance peak is neither very high nor very sharp. If, on the other hand, the probe extends only a small distance into the cavity, there is usually some difficulty in exciting the proper mode. For example, A of figure 124, shows the configuration of the electric field in a rectangular cavity resonating in the $TE_{0,1,1}$ mode, and B shows the configuration of the electric field set up by a probe. Because of this, the use of a probe usually is limited to applications where tight coupling is desired, such as to and from a waveguide, or from a directional coupler; or where the signal is set up by some other method and only a small amount of energy is to be taken out, such as from slotted sections of waveguide or coaxial line.



DESIRED ELECTRIC FIELD A

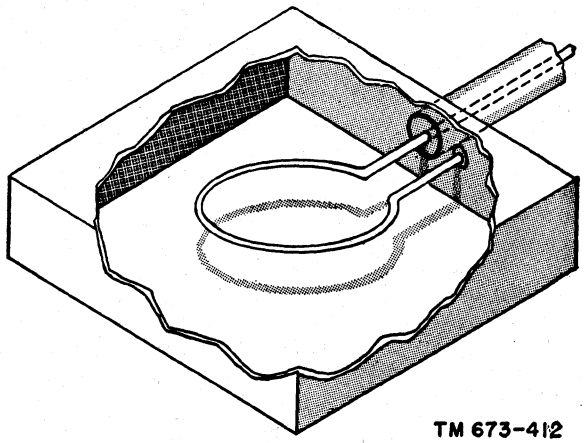


ELECTRIC FIELD SET UP BY PROBE B

TM 673-411

Figure 124. Effect of probe coupling on electric field.

b. Loop Coupling. Loop coupling is more common than any other method. Figure 125 shows a coupling loop used to excite the $TE_{0,1,1}$ mode in a rectangular resonator. The basic advantage of loop coupling lies in the fact that the coupling loop does not deform the electric field pattern to the same extent that a probe does. The degree of coupling is controlled by the size of the loop and sometimes, but less frequently, by the placement and orientation of the loop. If the loop is turned so that it lies in a plane parallel to the magnetic field of a given mode, it is not coupled at all to that mode, although there may be undesired coupling to other modes. Its operation in this respect is similar to that of the tuning paddle. One of the major uses of loop coupling is to take energy from the resonant cavity of a magnetron or a klystron, and it is used also to couple to and from the resonant cavities of some wavemeters (par. 86).



TM 673-412

Figure 125. Loop coupling to cavity resonator.

c. Slit Coupling. Slit coupling, which may involve small holes instead of slits, is used mostly in TR and ATR switches. Energy leaks into or out of the resonant cavity by means of the slits or holes in the same manner that it leaks from the main waveguide through two small holes into the auxiliary cavity (which is *not* resonant) of the

directional coupler discussed in paragraph 61. The principal merits of slit coupling lie in its mechanical simplicity and, in the case of TR and ATR switches, in the relatively smooth transition between the waveguide and the cavity when the coupling is tight.

Section II. APPLICATIONS OF CAVITY RESONATORS

86. Wavemeters

Most microwave wavemeters are essentially similar. A resonant cavity is coupled loosely to the source of signals and a detector is coupled to the cavity. The cavity then is adjusted to resonate at the operating frequency, resonance being indicated by a maximum reading at the detector, and the frequency is read from a scale on the tuning adjustment. Figure 126 shows the equivalent circuit of such a wavemeter, and figure 127 shows a cutaway view of the tunable cavity with its input and output coupling loops.

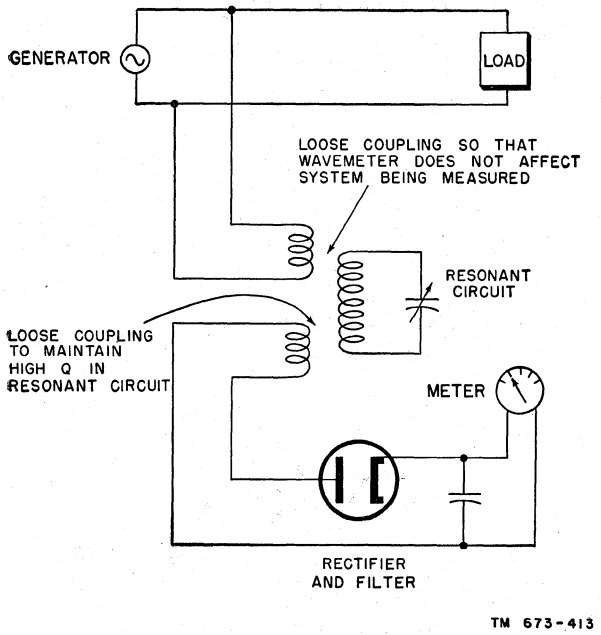


Figure 126. Equivalent circuit of wavemeter.

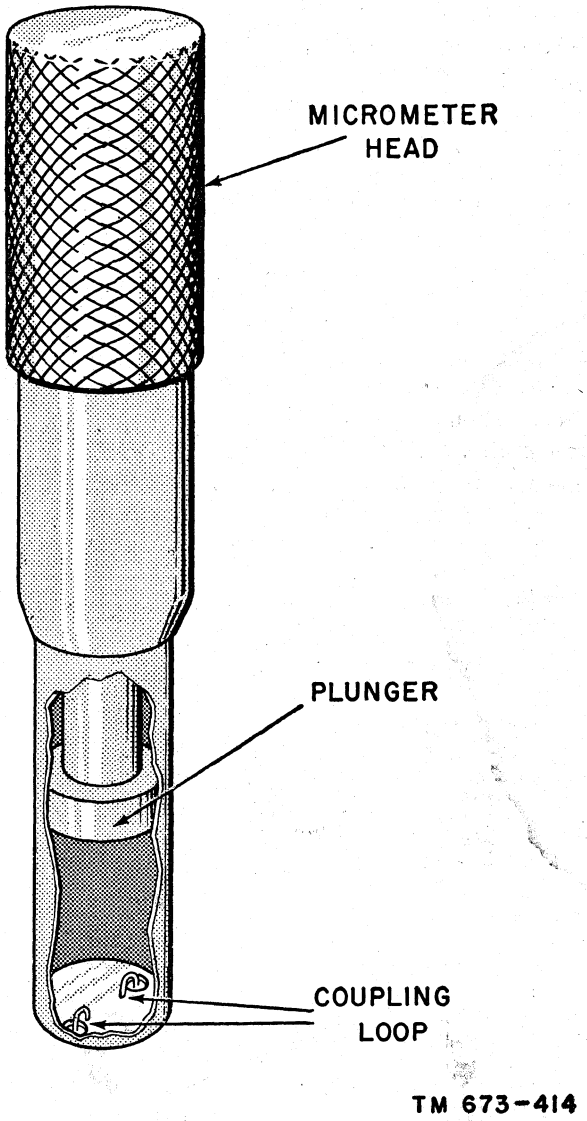


Figure 127. Wavemeter tunable cavity.

a. Sharpness of Resonance Peak. The impedance of a resonant circuit, whether it is a cavity resonator, a section of transmission line, or an L-C circuit, reaches a maximum at its resonant frequency (or at each of its characteristic frequencies) and falls off at frequencies above and below this. As the losses in the system are increased, the peak becomes lower and less sharp. It is customary to describe the sharpness of the resonance peak by means of a quantity, Q, which is defined as follows. Let Z_r be the impedance at the resonant frequency (or the characteristic frequency), f_r. At some frequency above f_r the impedance will have a magnitude equal to $Z_r \sqrt{1/2}$ or $.707Z_r$. Let this frequency be f_1. At some other frequency, below f_r, the impedance

will have the same magnitude. Let this frequency be f_2. Then, by definition,

$$Q=f_r/(f_1-f_2)$$

The Q of a cavity resonator is commonly several thousand.

b. Equivalent Circuit. In figure 126, the tuned circuit is coupled loosely to the transmission line in order to avoid disturbing the system the frequency of which is being measured. In addition, the detector is coupled very loosely to the resonant circuit in order to avoid lowering the Q. The detector shown consists of a conventional rectifier and filter circuit which feeds a meter, but, at microwave frequencies, the diode rectifier is replaced by a crystal rectifier and the filter capacitor is part of the capacitance of the crystal holder. In some cases, an amplifier may be inserted between the rectifier output and the meter. The resonant circuit is adjusted so that the rectifier output is maximized. When several frequencies are present, the meter will show a maximum as the resonant circuit is tuned through each one.

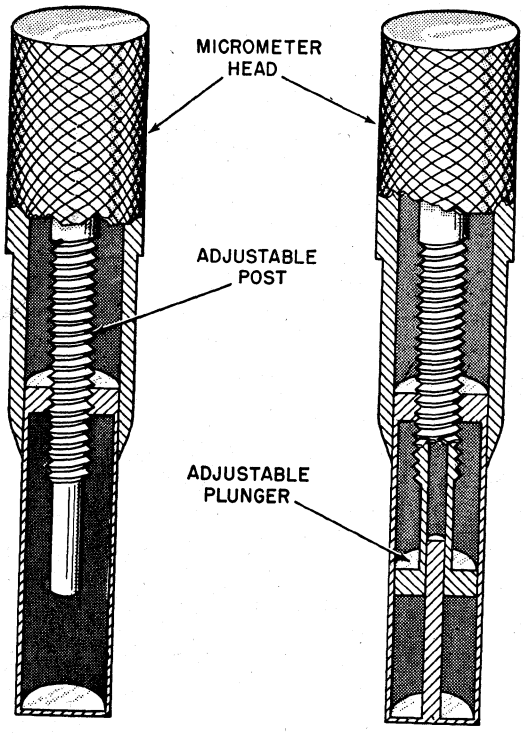


Figure 128. Wavemeter tunable cavities for low microwave frequencies.

c. Tunable Cavity. The tunable cavity shown in figure 127 consists of a cylindrical cavity of adjustable length operating in the $TE_{0,1,1}$ mode. The upper stop is moved by a conventional micrometer head until resonance is reached. The position of the plunger is read on a scale at the micrometer head and the frequency is determined from a calibration curve. A cavity of this sort is a precision instrument and must be treated with great care.

d. Other Forms of Cavities Used in Wavemeter. At the lower microwave frequencies, a simple cylindrical cavity becomes unduly large and various modified forms are used. Two of these types are shown in figure 128. Various forms of coupling are used, loop coupling being the most common.

e. Absorption Wavemeter. Another method of using a resonant cavity to determine frequency is shown in figure 129. Here a resonant cavity is coupled tightly to a waveguide so that it absorbs a substantial amount of energy when it is tuned to resonance. As the cavity is tuned to the operating frequency, the signal in the waveguide will drop.

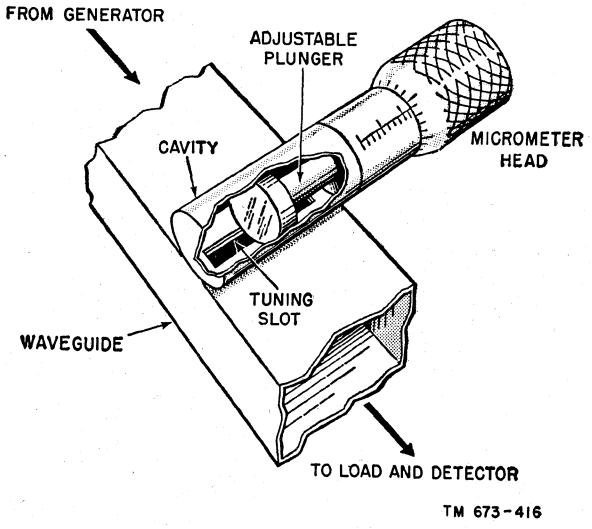


Figure 129. Resonant cavity used in absorption wavemeter.

87. Echo Boxes

An echo box is a piece of test equipment used to check the overall performance of a radar system. It consists of a tunable resonant cavity connected through a transmission system to a small antenna. The cavity is usually cylindrical, of adjustable length for tuning, and is operated in the $TE_{0,1,1}$ mode in order to obtain the highest possible Q.

a General Construction. An S-band echo box

(fig. 130) uses a small dipole antenna. This is connected through flexible coaxial cable to the resonant cavity where a small loop is used for coupling between the cable and the cavity. An X-band echo box generally uses a small dipole antenna at the center of a parabolic reflector. The cavity is mounted on the back of the reflector, and a short length of rigid coaxial line carries energy between the dipole and the coupling loop in the cavity. In most cases, the resonant cavity is the tuning knob in a micrometer head, and the echo box is provided with a calibration chart so that it can be used as a wavemeter. Occasionally, a second coupling device is added to the cavity. This coupling device generally is a probe which is part of a crystal-holder assembly similar to those discussed in chapter 3. The crystal rectifier output is fed to a meter so that the signal level in the cavity resonator can be measured.

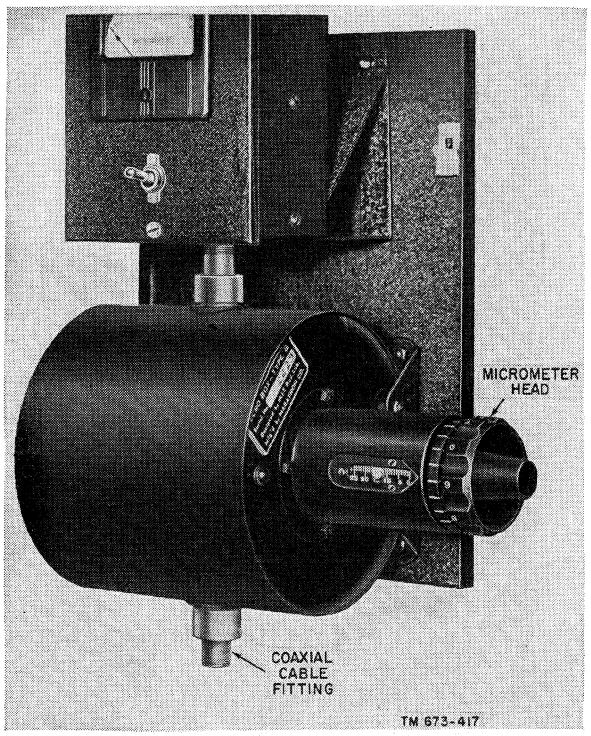


Figure 130. Echo box.

b. Behavior. If a cavity resonator is excited by an external signal, the level of the signal in the cavity will rise as shown in A of figure 131. Similarly, if the excitation is stopped suddenly, the signal level in the cavity will decay as in B. It is assumed that the exciting signal is at the resonant frequency of the cavity. The decay

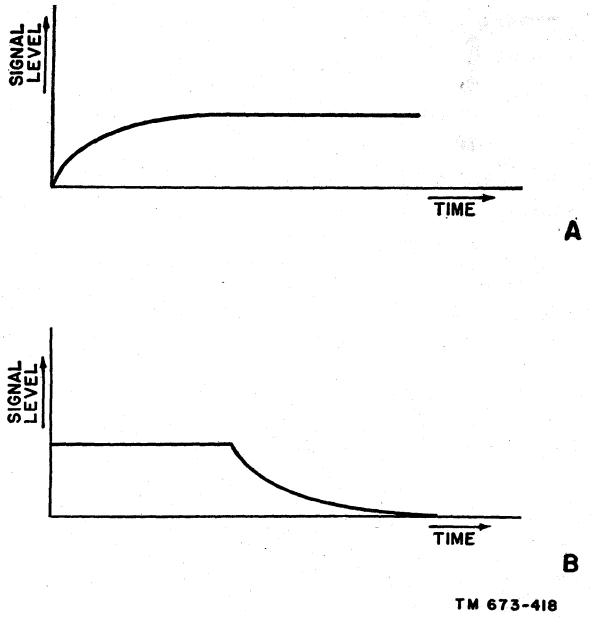


Figure 131. Rise and decay of signal in resonator.

time is related to the Q of the resonator, and the signal level will fall to .01 of its original value in $1.5Q$ cycles. Similarly, the rise time is related to Q. The signal level rises to .99 of its ultimate value (or reaches a point 1 percent short of its ultimate value) in $1.5Q$ cycles. The rise and decay times also can be expressed in terms of the *time constant* of the cavity, which is the time during which the signal level decays to about one-third (actually .368) of its original value, or during which it rises to about two-thirds (actually .632) of its ultimate value. In practice, the time constant of the cavity is so long that the signal level cannot rise to its ultimate value during the transmitter pulse. A graph of signal level against time is shown in figure 132.

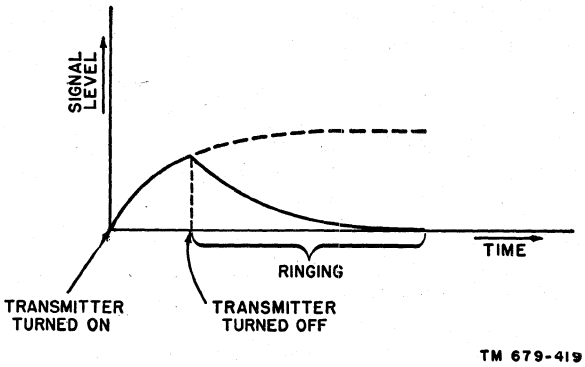


Figure 132. Signal level in echo box.

c. *Appearance of Echo Box Signal.*

(1) Figure 133 shows the appearance of the echo box signal on an A-scope, which displays receiver output versus range or time. During the transmitter pulse, the receiver is saturated and the output signal reaches its maximum level. Following the transmitter pulse, the receiver output results only from the signal picked up from the echo box (assuming that there are no short-range echoes). The output drops at first because the receiver sensitivity has been somewhat reduced as a result of saturation by the transmitter pulse, and also because the TR and ATR switches take some time to recover. The receiver, however, quickly recovers its normal sensitivity and the signal from the echo box is then sufficient to drive it to full output.

(2) After a short interval during which the receiver is saturated, the echo box signal drops off so far that the receiver output also drops. From this point, the signal from the echo box and the output from the receiver both decay until the receiver output is lost in the grass, or electrical noise, on the A-scope. The ringing time is, of course, the time interval between the end of the transmitter pulse and the point where the echo box signal is lost in the grass.

(3) It is customary to express the ringing time in whatever range units (yards or miles) are shown on the A-scope. The ringing time may be increased in three ways—

(*a*) Raising the transmitter output causes the echo box resonator to be excited at a higher level; it then takes longer for the echo box signal to decay to a point where the receiver output is lost in the grass.

(*b*) Increasing the duration of the transmitter pulse has the same effect, since the signal level in the resonator will rise further toward its ultimate value if the period of excitation is short compared to the time constant of the resonator. As before, a higher initial signal level will increase the time required for the signal to decay to the point where the receiver output is lost.

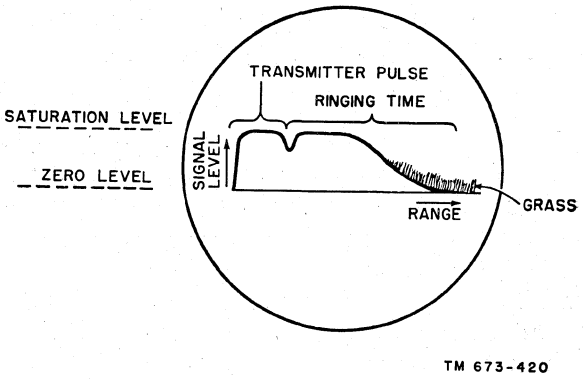


Figure 133. Echo box signal waveform on A-scope.

(*c*) An increase in receiver sensitivity will allow the receiver to *see* smaller signals, and, consequently, the period of decay will be longer to the lower critical level where the receiver output is lost.

(4) In this connection, it is necessary to distinguish between receiver sensitivity and receiver gain. Increasing the sensitivity raises the ratio of signal output to noise output for a given level of the input signal. Increasing the gain, however, raises both the signal output and the noise output without affecting their ratio. This will affect the appearance of the picture on the A-scope, but will not alter the ringing time. The appearances of the echo box signals on a B-scope, which displays range versus azimuth, and on a PPI scope (plan-position-indicator oscilloscope), which displays range versus direction, are shown in A and B, respectively, of figure 134. In each case, the receiver output controls the brightness of the spot on the scope. The ringing time always is taken from the greatest peak, which is obtained when the antenna is pointing at the echo box. When the antenna is pointed in any other direction, the ringing time decreases because the transmitter delivers a weaker signal to the echo box and also because the receiver output is reduced for a given signal *from* the echo box. The small peaks in the ringing time pattern result from *side lobes* in the antenna pattern. Although most of the energy is radiated in the main beam, a portion is radiated

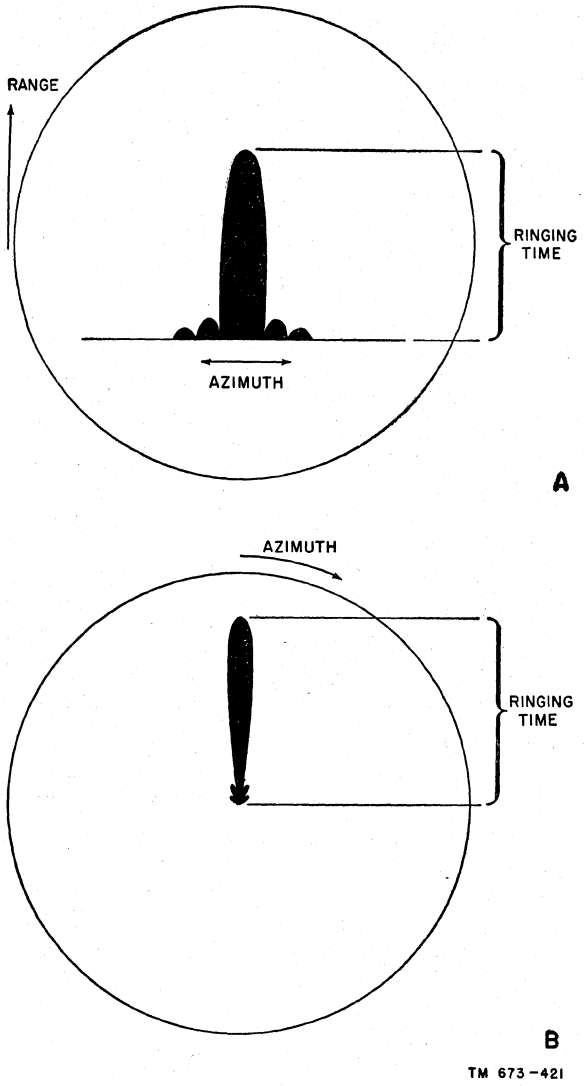


Figure 134. Echo box signals waveform on B-scope and PPI.

in directions near the beam. The behavior of the antenna for receiving is much the same. The strongest response from the receiver is obtained when the signal source lies in the main beam of the antenna, but signals can be picked up weakly from directions close to the beam. This is discussed more fully in chapter 7, which deals with microwave antennas.

d. Interpretation of Echo Box Test Results.

(1) In order to interpret the results of an echo box test, it is desirable to recall the expression developed for a given radar installation. Operating against a given target, the maximum range, unless limited by the horizon, is proportional to the fourth root of $P/_{p'}$, where P is the power output of the transmitter, and p' is the power input required by the receiver to develop a usable receiver output—an output signal that can be observed in spite of the electrical noise. This shows that an increase in the transmitter output or an increase in the receiver sensitivity, which is a decrease in p', will increase the maximum range against a given target and also will increase the ringing time with a given echo box. Therefore, any adjustment of the radar system which increases the ringing time, *except an increase in the duration of the transmitter pulse,* also will improve the system performance. The extent of the improvement can be calculated, either from the time constant of the resonator or from its Q, but the process is rather complicated and the calculation is not particularly useful. However, it is possible to estimate the improvement in performance when the echo box signal is displayed on an A-scope. First, note the range at which the echo box signal is just as strong as the noise signal on the scope. Then, after making an adjustment which increases the ringing time, note the new level of the echo box signal at the range where it was previously equal to the noise signal.

(2) The percentage increase in range against targets of a given type will be about half of the percentage increase in the ratio of echo box signal to noise signal. The initial ratio between the two signals may be considered 1, since the point of observation is chosen as the point where the signals are equal. If, after improvement, the echo box signal is 10 percent greater than the noise signal, the range on targets of a given type will be improved by about 5 percent. Although the echo box is one of the most valuable pieces of test equipment for tuning up a radar installation, it is dangerous to place much faith in comparisons between different ringing time tests unless the echo box always is placed in exactly the same position with the same orientation. Even then, any change in the tempera-

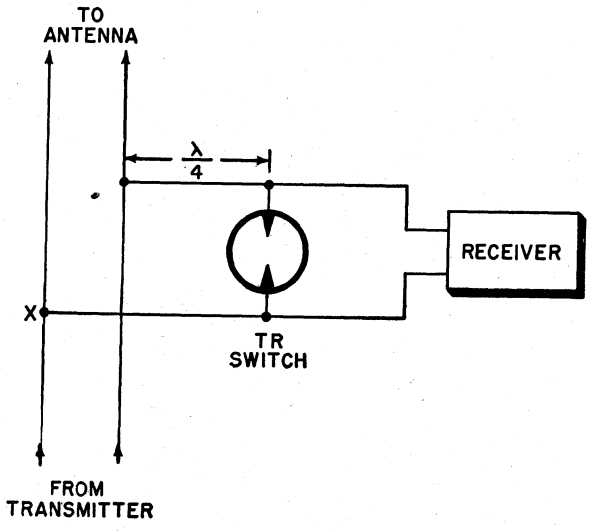
ture of the cavity resonator will alter its Q and, therefore, will change the ringing time, as will any change in the duration of the transmitted pulse. The system should be considered satisfactory if the ringing time is within about 20 percent of its usual value, and the echo box then should be used as an aid for the precise adjustment of the various tuning devices.

e. Procedure for Ringing Time Test. In view of the various different types of echo boxes, the procedure is more complicated to describe than to carry out. In fact, the ringing time test is particularly useful just because it is so easy. In any case, the transmitter is adjusted first. This problem is discussed in detail in chapter 5, and the reason for mentioning it here is to point out that the proper adjustment of the transmitter insures adequate output from the magnetron at the proper pulse duration. If the echo box is equipped with a detector, the next step is to tune it to the transmitter (by adjusting for maximum output from the detector). At this point, the transmitter frequency should be checked and the transmitter should be readjusted, if necessary, so that it operates on the proper frequency. (Unless the transmitter uses a tunable magnetron, its frequency cannot be altered and the frequency check is merely one step in the check for proper overall operation.) The receiver then is tuned to the echo box signal, with the automatic frequency control system turned off, and the various tuning devices in the transmission system are adjusted to obtain the longest ringing time. If the echo box is not equipped with a detector, the transmitter is adjusted first and then, by observing a local echo, the receiver is tuned to the transmitter, with the automatic frequency control system turned off. The echo box then is tuned to the transmitter by observing the ringing time and tuning the *echo box* for the longest ringing time. Thereafter, the tuning devices in the transmission system are adjusted to increase the ringing time.

88. TR and ATR Switches

In many early radar installations, the transmitter was connected to one antenna and the receiver was connected to another. Most modern radars, however, use the same antenna for transmission and reception. This eliminates the problem of pointing two antennas in the same direction and also effects a reduction in the size and weight of the antenna system. This improvement introduces two problems: First, it is necessary to protect the receiver from being damaged by the high-power transmitted signal; second, it is necessary to insure that none of the incoming signal goes to the transmitter, since diversion of any incoming energy from the receiver will lower the effective receiver sensitivity. The devices used to solve these problems are called TR and ATR switches. Essentially, they are nothing more than spark gaps, but at microwave frequencies they normally involve resonant cavities.

a. TR Switch Theory. The TR switch protects the receiver from damage by the high-power output of the transmitter (fig. 135), and insures that the bulk of the transmitter output goes to the antenna. Without the TR switch, a portion of the transmitter output would go directly to the receiver instead of being radiated from the antenna. When the transmitter is on, the signal level in the line is sufficiently high to cause the TR switch to arc and short-circuit the line to the receiver. The line to the receiver, when viewed from X, then appears as a quarter-wave shorted stub and has an infinite impedance. All of the transmitter output reaches the antenna except for the small amount necessary to maintain the arc across the TR switch. In addition, only a small amount of energy reaches the receiver, which thus is protected from damage. As soon as the transmitter is turned off, the arc goes out and the receiver is connected automatically to the antenna. With very high-powered transmitters, two spark gaps sometimes



TM 673-422

Figure 135. TR switch circuit.

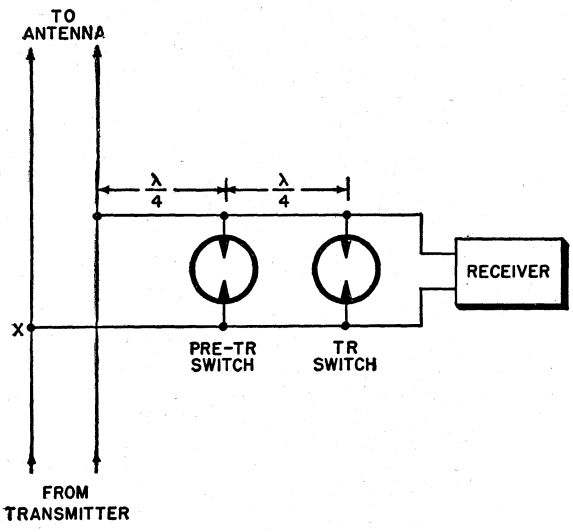


Figure 136. Pre-TR switch TR switch circuit

are used to afford added protection to the receiver. The spark gap nearer the transmission line is called a *pre-TR switch* and the one nearer the receiver is referred to as the TR switch (fig. 136).

b. ATR Switch Theory.

(1) The TR switch protects the receiver from damage and also prevents diversion of any significant portion of the transmitter output from the antenna. It is not, however, effective during reception and cannot prevent the incoming signals from being partly wasted at the transmitter. One method of preventing this waste of incoming energy, shown in figure 137, is to use a line-stretcher between junction X and the transmitter. The length of the line between X and the transmitter is adjusted, by use of the line-stretcher, so that the impedance at X looking toward the transmitter, with the transmitter turned off, is infinite. All of the incoming energy then goes to the receiver. Sometimes the line between X and the transmitter is preplumbed and the line-stretcher is omitted.

(2) A more common method of insuring that all the incoming energy reaches the receiver is by use of an ATR switch as shown in figure 138. When the transmitter is on, both the TR switch and the ATR switch arc. The impedance looking from X toward the TR switch is then infinite, since this branch of the line looks like a quarter-wave shorted stub. For the same reason, the impedance at Y looking toward the ATR switch is also infinite. All of the transmitter output, except that necessary to maintain the arcs, reaches the antenna, and the receiver is protected. When the transmitter is turned off, the arcs go out and the receiver is connected to the transmission line, as explained above. In addition, the impedance at point Y looking toward the ATR switch is now zero, since this branch is now a half-wave shorted stub. The line then is short-circuited effectively at Y and the impedance at X looking toward the transmitter is that of a shorted quarter-wave stub, which is infinite, and all of the incoming energy goes to the receiver. The effect of the combined switch system is as follows: The receiver is protected from damage—

(*a*) When the transmitter is on, the receiver is removed from the transmission line and all of the transmitter output goes to the antenna.

(*b*) When the transmitter is off, the receiver is connected to the transmission line and the transmitter is removed from the line so that all of the incoming energy goes to the receiver.

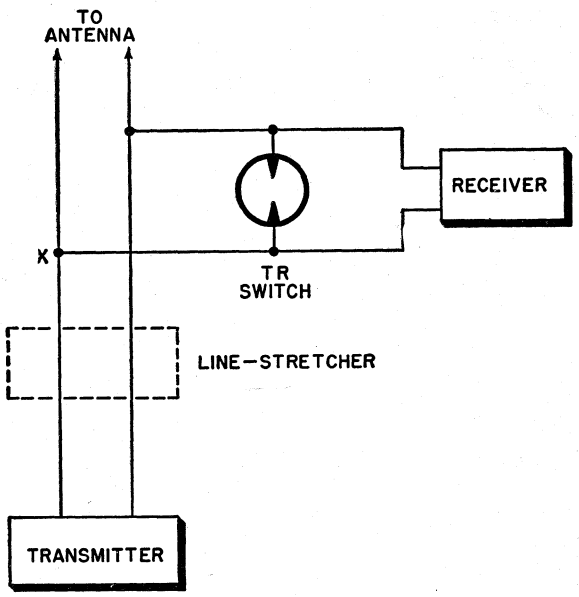


Figure 137. Line-stretcher in TR switch circuit.

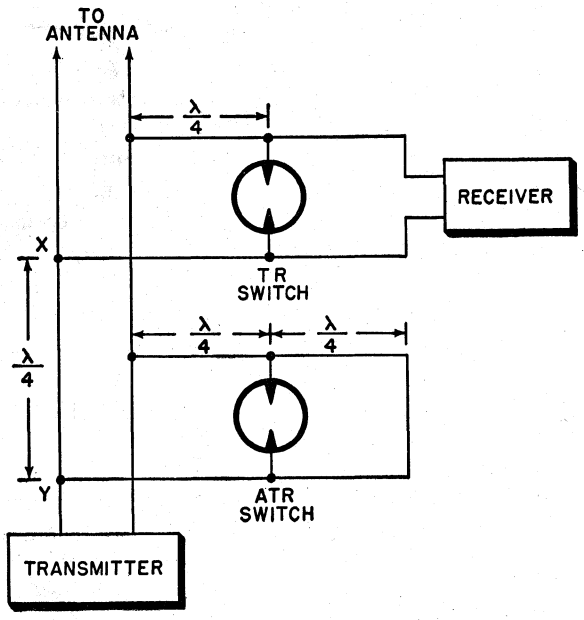


Figure 138. ATR switch circuit.

c. Construction. The TR and ATR switches are essentially similar in construction. Each consists of a resonant cavity with a pair of electrodes across which an arc is struck when the transmitter is turned on. Coupling to and from the cavity is by means of loops when the switch is used with coaxial line, and by means of windows (which are large slits) when the switch is used with waveguide. The cavity normally is equipped with both tuning paddles and tuning plugs, the paddles being used for coarse tuning, and the plugs for fine tuning. In order that the arc may be easily formed, the electrodes are contained in a glass bulb filled with gas under low pressure. The complete assembly (bulb with electrodes) is called a TR tube. One of the electrodes frequently is mounted on a flexible diaphragm which forms part of the wall of the bulb, and the spark gap spacing then can be adjusted by deforming the diaphragm. It is customary to provide a third electrode, called a *keep-alive* electrode. A voltage, called the keep-alive voltage, is applied to this electrode and a steady arc is maintained between one side of the spark gap and the keep-alive electrode. This subsidiary arc permits the main arc to form more quickly when the transmitter is turned on. A, of figure 139, shows an S-band TR tube, and B shows an X-band TR tube. It is necessary to provide a current-limiting resistor in the keep-alive line, and this must be placed close to the TR tube to

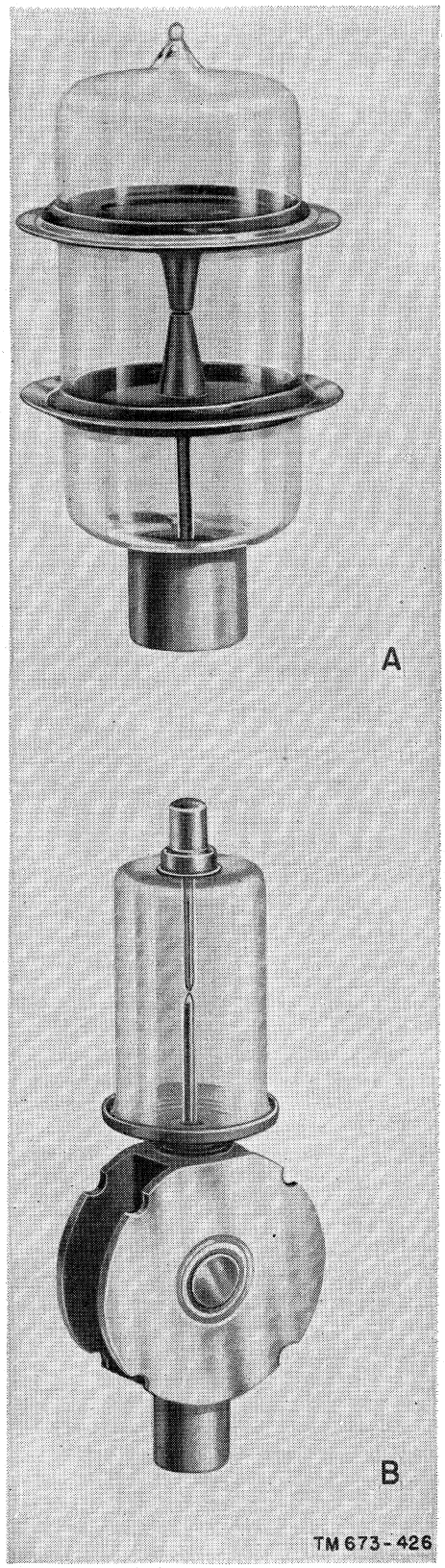


Figure 139. TR tubes.

avoid oscillation of the keep-alive arc, which is similar to the oscillation of a thyratron oscillator, and consequent damage to the receiver when the TR switch fails to arc. Figure 140 shows a diagrammatic representation of the complete TR–ATR switching system, and figure 141 shows an X-band TR–ATR assembly.

d. Application to Radar System. The application of TR and ATR switches in a radar system is shown in figure 142. There is considerable variation in systems of this sort. For instance, one or more of the tuners may be omitted if the system is preplumbed, or additional tuners may be added if the transmission line is long and if it has many bends or joints. Directional couplers also may be added at strategic points to aid in the tuning.

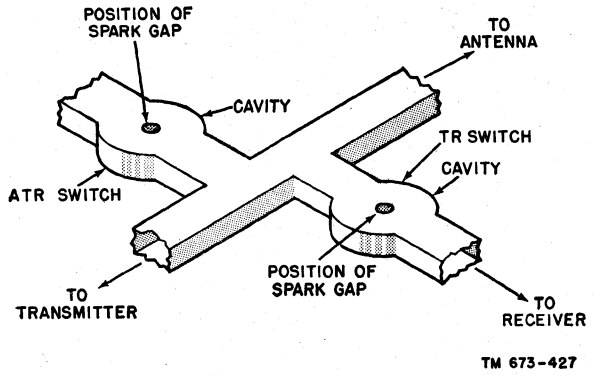


Figure 140. TR–ATR switch system.

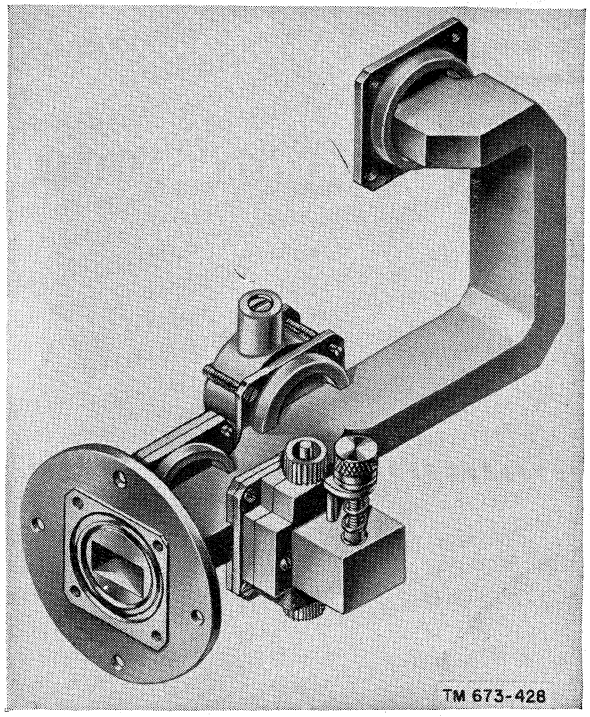


Figure 141. X-band TR–ATR switch assembly.

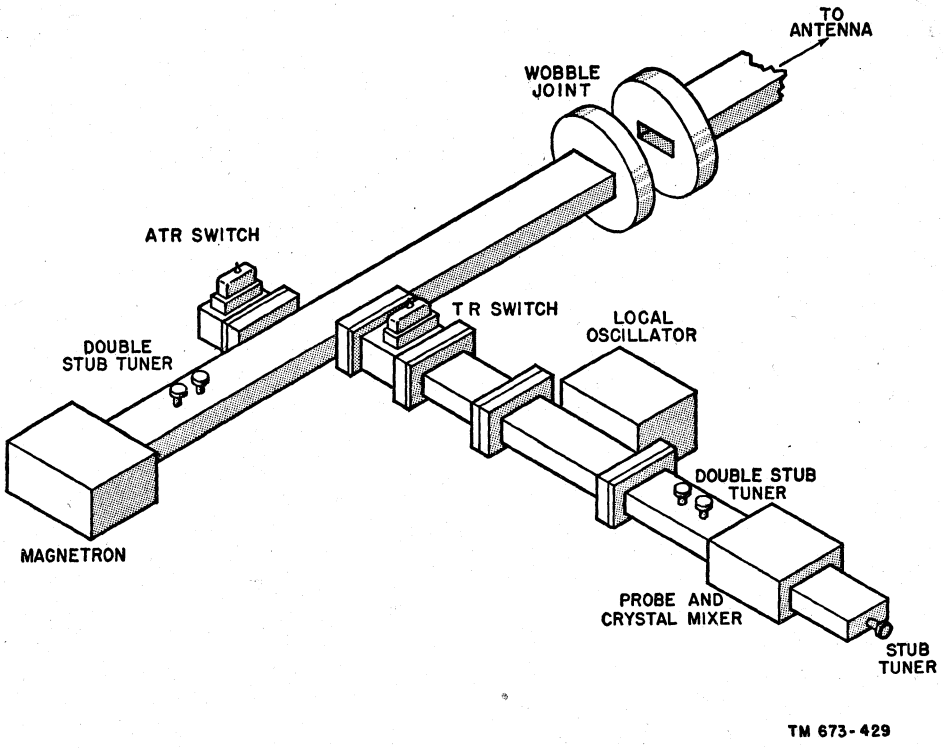


Figure 142. Application of TR and ATR switches to radar system.

89. Summary

a. Resonant cavities, or cavity resonators, are used at microwave frequencies in place of the *L–C* circuits, and transmission line stubs are used at lower frequencies. A cavity resonator is a hollow chamber with conducting walls.

b. Although a cavity resonator may be of any shape, it always can be considered as a distorted version of a simple resonator consisting of a length of waveguide terminated by conducting walls. With a rectangular resonator, there are infinitely many modes of operation, since either TE or TM signals may be propagated from end to end inside the cavity, from side to side, or between the top and bottom of the cavity. Each of the modes has a distinct field pattern, and the cavity may be resonated at any of the modes or at several modes simultaneously. There are an infinite number of resonant frequencies and, in general, several modes corresponding to each one.

c. In the neighborhood of one of its resonant frequencies, sometimes called characteristic frequencies, a resonator behaves like a tuned circuit. Its input impedance rises to a maximum at resonance and, like a tuned network, it is most effective in transferring energy, from its input terminals to its output terminals, at resonance.

d. Cavity resonators are used in microwave wavemeters, in echo boxes, and in TR and ATR switches. They are included also, as part of the tube structure, in microwave oscillator tubes.

90. Review Questions

a. What is a cavity resonator?

b. What device may be used to take energy from a cavity resonator?

c. What is the impedance of an *L–C* circuit at the resonant frequency of the circuit?

d. In a resonant transmission line section, which mode of resonance has greatest wavelength?

e. Name three methods of tuning a cavity resonator.

f. Does a tuning paddle parallel to the direction of the magnetic field have an effect on the field?

g. What is the position of a tuning paddle when the magnetic field distortion is greatest?

h. Name three ways of coupling to a cavity resonator.

i. When a probe coupling is used in a cavity resonator, is the amount of energy to be taken out large or small?

j. Which is the most common method of coupling to a cavity resonator?

k. What is the purpose of an echo box?

l. In general, an adjustment of a radar system that increases ringing time (when adjusting with an echo box) improves the system performance. What is the exception?

m. What are the functions of TR and ATR switches?

n. What does the TR switch do when the transmitter is on?

CHAPTER 5

MICROWAVE OSCILLATORS

Section I. OSCILLATORS IN GENERAL

91. Klystrons and Magnetrons

The preceding chapters of this manual have demonstrated the marked differences between microwave circuit techniques and those used at lower frequencies. A conventional transmission line, for example, often is replaced by a waveguide at microwave frequencies, and a cavity resonator takes the place of the more familiar coil-capacitor combination. In view of these differences, it is not surprising that microwave oscillators have little in common with the usual vacuum-tube oscillators used in lower frequencies. There is, at present, no satisfactory method of amplifying microwave signals at high-power levels, and microwave oscillators must, therefore, be connected directly to the loads they feed. Two distinct types of microwave oscillators have been developed for low-power and high-power applications. A *klystron* is used at microwave frequencies where a simple oscillator (or an oscillator and a buffer) would be used at lower frequencies. The average power output of a klystron is, in general, something less than 1 watt, and klystrons are used in signal generators and other test equipment, in low-power transmitters, and as local oscillators in microwave superheterodyne receivers. A *magnetron*, on the other hand, is used at microwave frequencies where an oscillator and a power amplifier would be used at lower frequencies. The usual microwave transmitter with medium- or high-power output uses a magnetron oscillator. The complete transmitter consists of the magnetron, a microwave transmission system (generally using waveguide), an antenna, a power supply, and a modulator to provide amplitude, frequency, or pulse modulation. The magnetron thus takes the place of a conventional low-frequency oscillator, buffer, frequency-multiplier, voltage amplifier, driver, and power amplifier. Both types of microwave oscillators are vacuum tubes of a rather special sort called *velocity modulated* tubes. The internal elements of a velocity-modulated tube acts on the electron stream by changing the average *velocity* of the electrons; those of a conventional vacuum tube control the *number* of electrons in the stream. The theory of velocity modulation is discussed in paragraphs 101–105, which covers klystron oscillators.

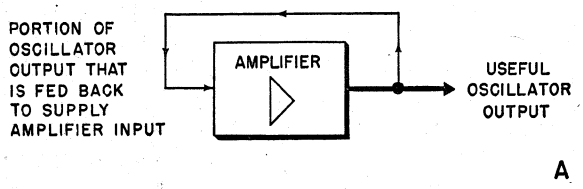
92. Mechanism of Oscillation

Since each type of microwave oscillator makes use of velocity modulation—a technique ordinarily not used at lower frequencies—a study of either klystrons or magnetrons must include two distinct subjects. First, how the tubes work, how they are constructed, and how the tube elements and the electron stream act on each other; second, why the tubes oscillate. Therefore, a general understanding of the various mechanisms of oscillation is necessary, and a review of oscillator theory will make it easier to understand how klystrons and magnetrons generate microwave signals. An explanation of conventional vacuum tube oscillators and their limitations will cover different types of vacuum tubes required for operation at microwave frequencies. The remainder of this section is, accordingly, devoted to a discussion of three basic types of oscillators: feedback, relaxation, and ringing oscillators. Paragraphs 96–100 describe some of the difficulties encountered when conventional oscillators are operated at high frequencies.

93. Feedback Oscillators

Feedback oscillators, which are quite simple, are probably more common than any other type in the field of electronics. A feedback oscillator is noth-

ing more than an amplifier that supplies its own input. A of figure 143 shows a block diagram of the basic feedback amplifier loop. A small portion of the amplifier output is fed back to supply the amplifier input, and the remaining portion is delivered to an external load as the output of the oscillator. B shows a type of feedback oscillator often found when an inexperienced technician sets up a public address system. Although this is a highly undesirable system in practice, it is a good illustration of a feedback oscillator. The microphone, electronic amplifier, and loudspeaker form an acoustic amplifier, which receives an input of low-level sound and delivers an output of sound at a higher level. Here, the feedback signal is in the form of sound. Following the terminology used in A, the signal delivered by the loudspeaker is labeled *useful output*, although one familiar with the problems of public address systems may find it difficult to say what it can be used for when the system is oscillating.



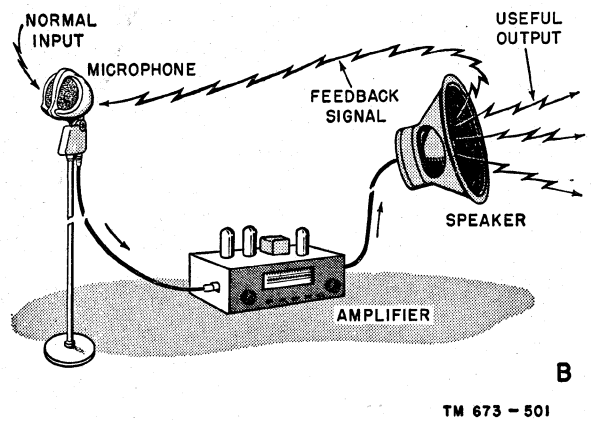


Figure 143. Basic loop of feedback oscillator.

a. Loop Gain. The system shown in A of figure 144, consists of an amplifier having a gain of a and a feedback path having a gain of b (or an attenuation of $1/b$). Since the operation of an ordinary low-power electronic amplifier is described most conveniently in terms of input and output voltage levels, it is assumed that a and b are voltage gains, rather than power gains. With

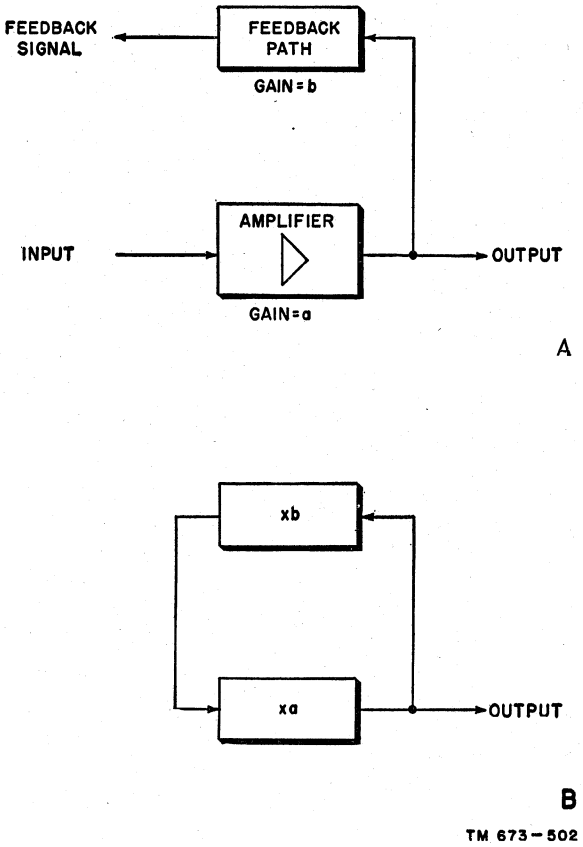


Figure 144. Concept of loop gain.

a system of this sort, the output is a times the input, and the feedback signal is ab times the input. If the product, a, which is the *loop gain*, is exactly equal to 1, the loop may be closed, as shown in B, and the feedback signal will supply exactly the amplifier input that is required for steady oscillation. Suppose, for example, that a 1-volt signal from an external source is connected to the amplifier input in the system shown in B. This will develop a feedback voltage of exactly 1 volt, having the same phase as the amplifier input, and the feedback signal may be substituted for the external signal, by closing the loop, as in the illustration, with no change in the signal levels at various parts of the system. If the loop gain is less than 1, the closed loop system is unable to supply itself with an adequate input signal, and the oscillation will die away, as shown in A of figure 145. If, on the other hand, the loop gain is greater than 1, the feedback signal is greater than the required input for steady oscillation and the amplitude of oscillation will increase, so long as the loop gain exceeds 1, as in B.

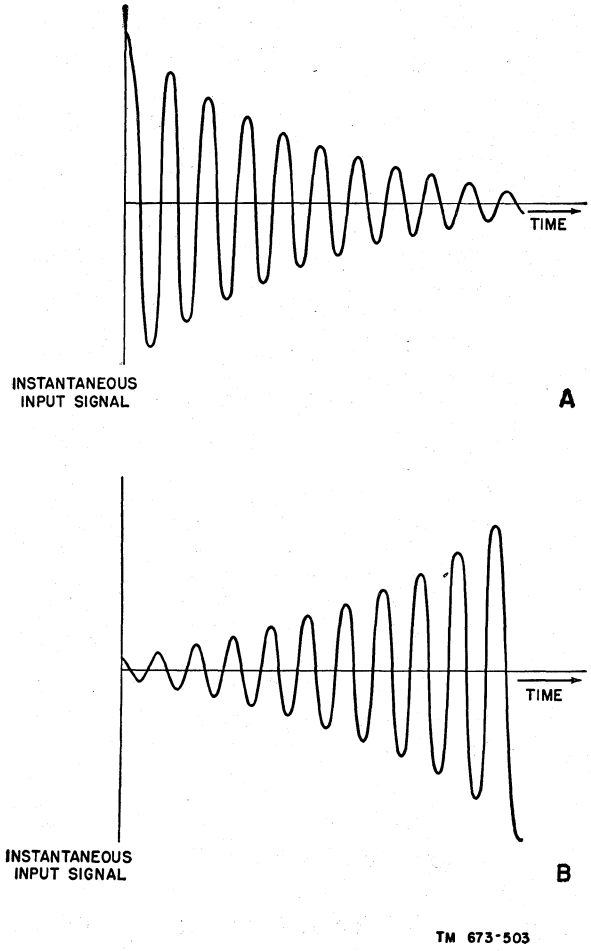


Figure 145. Effect of changing loop gain.

b. Establishing Steady Oscillation. Although this discussion explains how oscillations can sustain themselves, grow, or die away in a feedback oscillator, nothing has been said about how oscillations start or why the oscillator settles on a particular frequency. The key to this lies in *saturation* of the amplifier portion of the feedback oscillator. Saturation, which should be familiar to every electronics man, is the reduction in the gain of an amplifier when the signal level increases beyond some critical point. A of figure 146, shows the output of an amplifier for various input levels. For small inputs, less than 1 volt, the output is five times the input, and the gain, shown in B, is 5. Although the output continues to rise as the input signal is increased, the gain (the ratio of output to input) decreases. For an input of 2 volts, the output is only 6 volts and the gain is reduced to 3. By the time the input reaches 5 volts, the gain has fallen to 1.5; for an input of 8.25 volts, the gain is only 1. With still higher input signals, the output continues to rise slightly, approaching a saturation level of 8.5 volts, whereas the gain continues to decrease, reaching .5 at a 17-volt input level. This set of figures has been chosen for the purpose of making a clear presentation of saturation. In practical amplifiers, however, the gain for small signals is likely to be 50 or more and may be as great as several hundred. With a clear picture of saturation in mind, it is easy to see how the oscillator behaves. With no signal in the amplifier, the loop gain (*ab*) is high. Suppose, then, that there is some disturbance in the circuit—probably the result of tube noise. This disturbance may be thought of as composed of a number of distinct frequencies; for some of these frequencies, the loop gain will be high and with the proper phase for the amplifier to supply its own input. Oscillations will *start* and *grow rapidly* at all of these frequencies. As soon as the signal level rises in the amplifier, the gain (and consequently the loop gain) will be reduced. As the oscillations grow, the loop gain will become smaller until it is less than 1 for some of the frequencies of oscillations (for which the loop

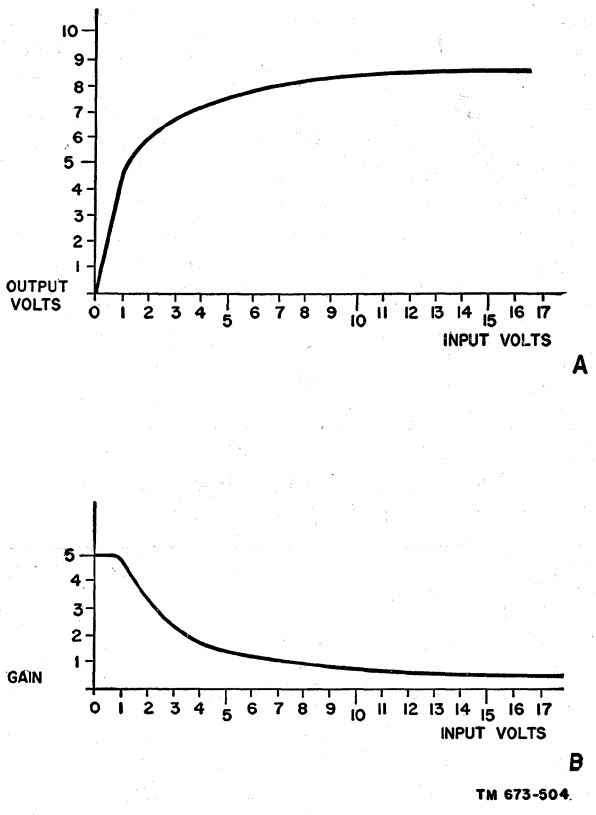


Figure 146. Amplifier saturation.

gain for small signals was greater than 1). The oscillations at these frequencies then die away quite rapidly. The process continues as long as the over-all signal level increases (as long as the loop gain remains greater than 1 for one or more frequencies). In a well designed oscillator, all signals but those at one frequency die away shortly after the oscillator is turned on. With poor design, there may be two or more frequencies at which the oscillator operates, and in some types of oscillator it is extremely difficult to confine the operation to a single frequency. The oscillator, therefore, reaches stable operation at a single frequency with a signal level for which the amplifier gain is $1/b$ (so that the loop gain is 1). All other frequencies are suppressed because ab is less than 1 for these frequencies. The signal level is stable, since any increase in signal level causes a reduction in loop gain and a consequent dying away of the oscillation. Similarly, any reduction in level causes an increase in loop gain and a growth of the oscillation.

94. Relaxation Oscillators

A relaxation oscillator is probably easier to understand than any other type. It consists of a system, of almost any sort, that tends to *relax* toward a preferred state. Before the preferred state is reached, a device is triggered that pushes the system *away* from the preferred state, after which the system starts once more to relax. The process is something like what would happen if a tired man were to sit on a tack, provided that for some reason he attempted to sit on the tack again and again. The period of a relaxation oscillator is the time interval between successive triggering of the displacement device. This interval depends on how far the system is displaced from its preferred state and on how rapidly it relaxes. The time required for displacement *from* the preferred state is usually negligible.

a. Thyratron Oscillator. The best known relaxation oscillator is undoubtedly the thyratron sweep circuit used in cathode-ray oscilloscopes of moderate speed. One version of this oscillator is shown in A of figure 147. The preferred state is, of course, the condition in which the firing capacitor is charged fully so that no current flows through the charging resistor (or through the bias network). Before this state is reached, however, the thyratron fires and discharges the firing capacitor almost to zero. Then the capacitor charges

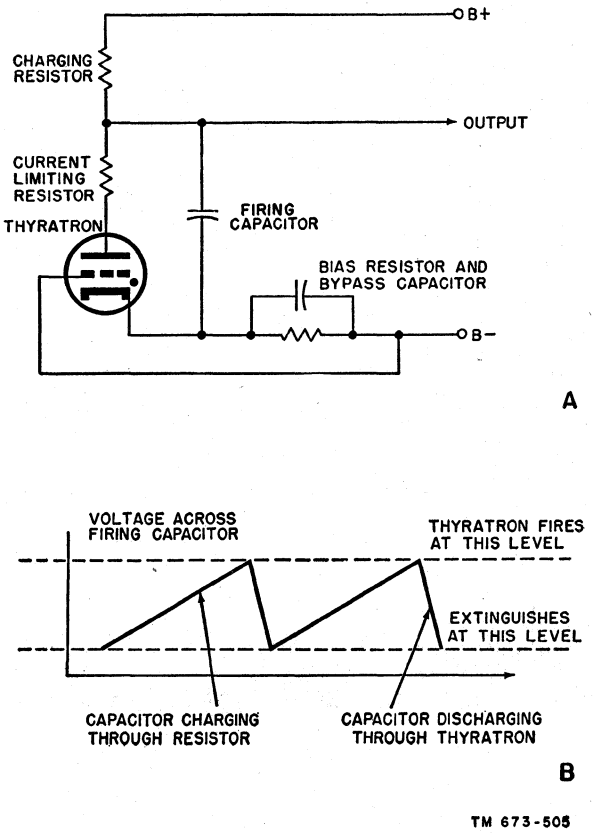


Figure 147. Circuit of relaxation oscillator with thyratron.

through the charging resistor until the voltage across it is once more sufficient to cause the thyratron to fire. The waveform of the voltage across the firing capacitor is shown in B. The charging rate—the rate at which the system approaches its preferred state—is set by the time constant of the firing capacitor and the charging resistor. An interesting variation of this circuit is shown in A of figure 148. Here the firing capacitor discharges through the discharge resistor until the thyratron cathode drops enough for the thyratron to fire. The capacitor then is recharged by the thyratron, which extinguishes itself when the anode-cathode voltage falls below a critical value. Since the thyratron fires at a critical value of grid-cathode voltage (whereas the earlier circuit fires at a critical value of anode-cathode voltage), the firing point can be adjusted easily by altering the grid bias. This, of course, changes the instant at which the system is pushed *away* from its preferred state, and so controls the frequency of oscillation. A modification of this circuit that provides a more linear charging rate is shown in B. The charging resistor is replaced, here, by a

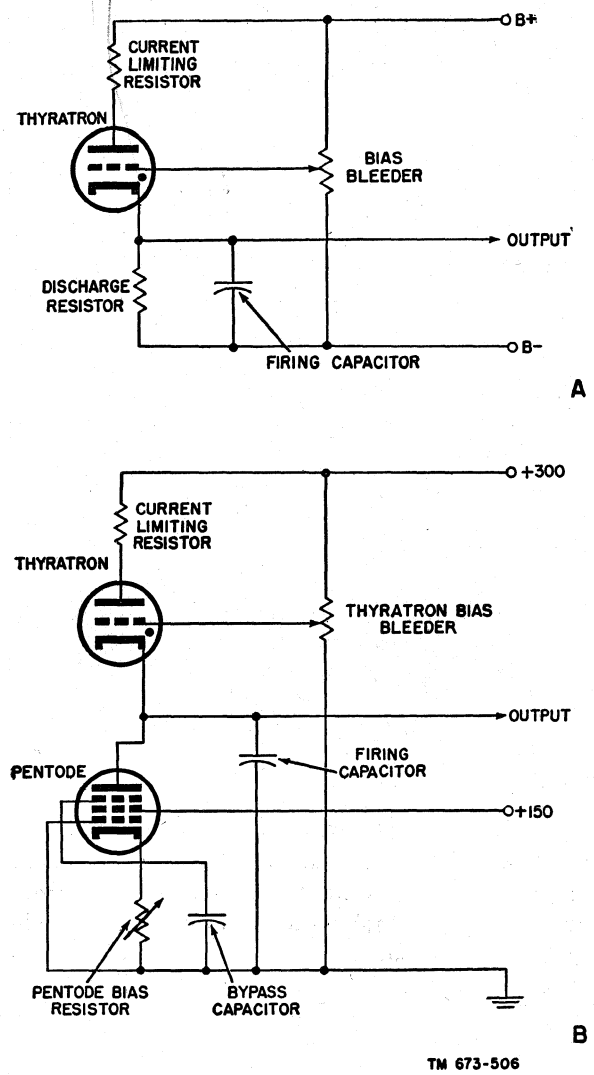


Figure 148. Modified thyratron oscillator circuit.

pentode, which functions as a constant-current device. Coarse frequency control now is achieved by adjustment of the thyratron bias, and fine frequency control is obtained through adjustment of the pentode bias resistor.

b. Multivibrator. A conventional multivibrator (fig. 149) may be thought of as a push-pull relaxation oscillator. The preferred state of each tube is the conducting state, and as one tube reaches its preferred state and starts to conduct, it drives the grid of the other tube well past cut-off. It is desirable to examine the way in which the multivibrator starts to oscillate. Suppose that the two tubes are in the circuit as shown, with the single exception that the two coupling capacitors are omitted. The current through each tube then assumes a steady value which is determined by the anode and cathode resistors, and the fact that the grid is returned to B-minus. If, now, the coupling capacitors are connected, there is no change in the system, at first. Any variation in the grid voltage of either tube, however, will throw the system into oscillation. Suppose that the grid of tube 1 moves slightly in the positive direction as a result of tube noise in either one of the tubes. This momentary positive signal is amplified in tube 1 and drives the grid of tube 2 in the negative direction. Tube 2, in turn, amplifies the signal at its grid and returns a large positive signal to the grid of tube 1. The process is cumulative, and proceeds with increasing speed—the grid of tube 1 moving in the positive direction while the grid of tube 2 moves in the negative direction. This continues until the two-tube amplification can proceed no farther—usually as a result of cut-off in the tube the grid of which is driven in the negative direction; the process stops when tube 2 is cut off, with the grid of tube 1 well above its normal level. This condition is not permanent, however, because the coupling condenser that feeds the grid of tube 2 begins to shift its charge, discharging through the grid resistor of tube 2 and the anode resistor of tube 1. Tube 2 then starts to conduct, and the two-tube amplification process lifts the grid back into the positive region and drives the grid of tube 1 well past cut-off, as explained above. The behavior of the system as it goes into oscillation is a little like that of a pencil that has been balanced on end. Any disturbance will throw it out of its condition of unstable equilibrium and into oscillation. The circuit can be made to have greater amplification by connecting the cathodes of the two tubes to B-minus through a single cathode resistor, and this is frequently done.

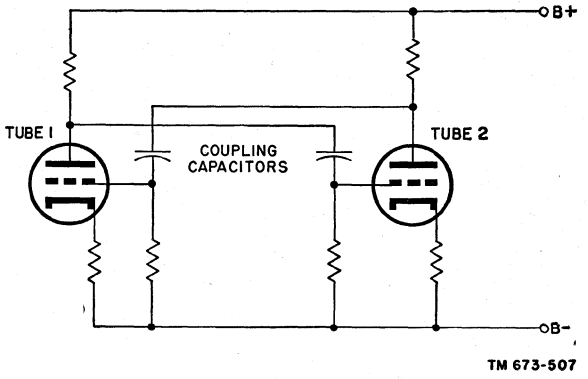


Figure 149. Multivibrator circuit.

c. Univibrator. The circuit known as a *univibrator* relaxes toward a preferred position and does *not* trigger a device which displaces it. This circuit, which is sometimes called a one-shot multivibrator in the United States and a Kipp relay in Great Britain, is shown in figure 150. In the preferred state of the system, tube 2 conducts, since its grid is returned to its cathode and the tube has a grid bias of zero. The bias bleeder for tube 1 is adjusted so that tube 1 is just a little past cut-off. In the quiescent state, then, tube 2 draws current and tube 1 does not. The circuit is triggered by a positive pulse at the input terminal, and the input need be only large enough to cause tube 1 to draw current. This signal is amplified in tube 1 and appears as a negative pulse at the grid of tube 2. The resulting drop in the current of tube 2 causes the cathodes of both tubes to move in the negative direction, and the negative signal at the cathode of tube 1 causes precisely the same result as the original positive signal at the grid of tube 1. As in the multivibrator, the process continues until tube 2 is cut off. At this point, tube 1 draws current and tube 2 does not. The coupling capacitor that feeds the grid of tube 2 then discharges, and as soon as tube 2 commences to conduct, the initial process is reversed and tube 1 again is cut off. At this point, however, the system rests until another input pulse is received. The duration of the unstable state (with tube 1 conducting and tube 2 cut off) is determined by the resistor and capacitor connected to the grid of tube 2.

d. Relay Relaxation Oscillator. It is necessary to emphasize the fact that an oscillator need not be electronic. In the following paragraph on ringing oscillators, for instance, an old-fashioned pendulum clock is mentioned as an example, and an electrical but *nonelectronic* oscillator is shown in figure 151. The latter consists of the usual capacitor and charging resistor together with a relay. In operation, the capacitor charges through the resistor until the voltage across it is sufficient to pull in the relay. When the relay is actuated, one set of contacts is used to short the capacitor, after which the relay drops open and the capacitor starts once more toward the preferred position of full charge.

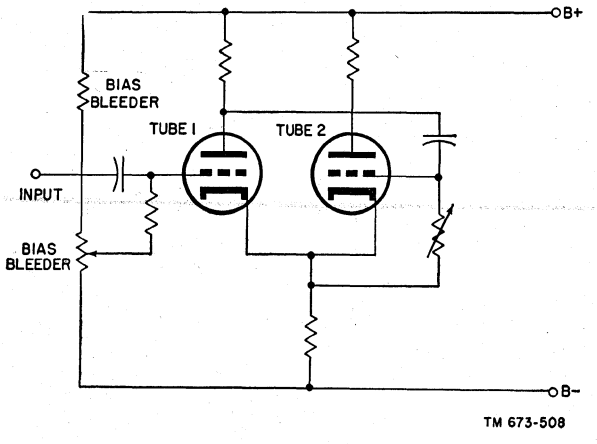


Figure 150. Univibrator circuit.

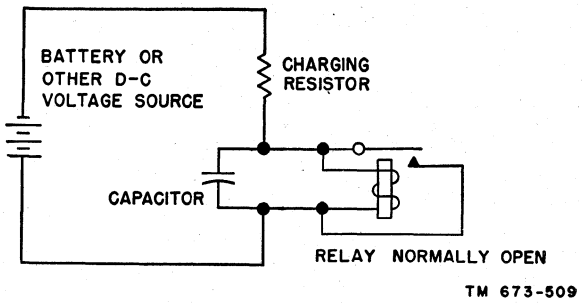


Figure 151. Circuit of relaxation oscillator, using relay.

95. Ringing Oscillators

Another type of oscillator is called a ringing oscillator. This consists of a resonant system, which may be a pendulum or a cavity resonator, together with a mechanism of some sort for feeding energy to the resonator. It has been explained (ch. 4) that a cavity resonator will ring at its own resonant frequency whenever it is shock-excited, and the same thing is true of any other resonator. If, then, matters can be arranged so that the resonator periodically is given a small boost *at the proper time,* it will continue to ring and the combination of resonator and boosting mechanism legitimately may be considered an oscillator. One good example of a ringing oscillator is the pendulum on an old-fashioned clock. Left to itself, the pendulum would swing from side to side at its resonant frequency but the amplitude of its swings would decrease continually and finally it would come to rest. The clock mechanism, however, provides a gentle push to the pendulum at each end of its swing, and it goes on oscillating indefinitely so long as someone remembers to wind the clock. Another example, of course, is the balance wheel and escapement mechanism on a watch or a modern nonelectric clock. This, however, may not be as familiar as the pendulum. Still another

example of a ringing oscillator is a small boy in a swing. If he *pumps* at the right time during each swinging cycle, he will oscillate indefinitely, or until he gets hungry and goes home for supper. A good example of the ringing oscillator in the field of conventional electronics is the tri-tet harmonic oscillator shown in figure 152. This consists of a pentode that operates like a combination of a triode oscillator and a buffer stage. The cathode, control grid, and screen grid of the tube are connected as a triode oscillator of one sort or another; the illustration shows them connected as a conventional crystal oscillator. The tuned circuit in the plate line then is set to resonate at the desired harmonic of the oscillator frequency. Its behavior is very much like that of a clock pendulum, except that it gets a boost in each direction only once during each cycle of the oscillator. Since the resonator is boosted one way when the oscillator current is at a maximum and the other way 1 half-cycle later when the oscillator current is at a minimum, the system works best as a generator of odd harmonics. It should be noted that all frequency multipliers are not ringing oscillators. The conventional push-push doubler, for example, shown in figure 153, operates with the inputs to the two tubes in push-pull and the two outputs in parallel. If the tubes are operated class B, the output looks like that of a full-wave rectifier and it has a very large second harmonic content. The output tank then serves as a filter rather than as a resonator, although the distinction is slight, and selects a second harmonic signal, or other even harmonic signal, which is already present. The harmonic tank in the tri-tet oscillator, however, does not receive any significant amount of harmonic signal and actually generates the harmonic output by ringing.

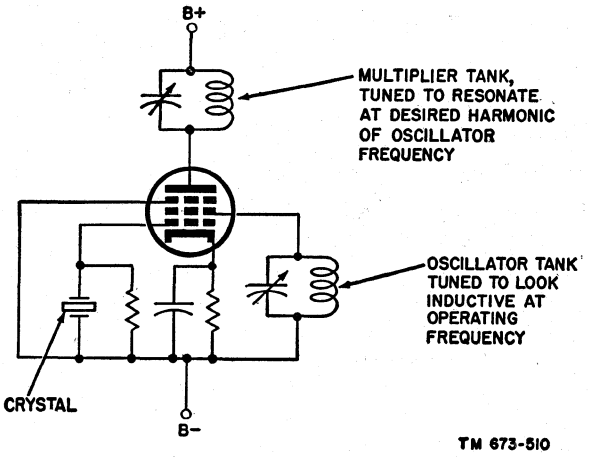

B+
MULTIPLIER TANK, TUNED TO RESONATE AT DESIRED HARMONIC OF OSCILLATOR FREQUENCY
OSCILLATOR TANK TUNED TO LOOK INDUCTIVE AT OPERATING FREQUENCY
CRYSTAL
B-
TM 673-510

Figure 152. Tri-tet harmonic oscillator circuit.

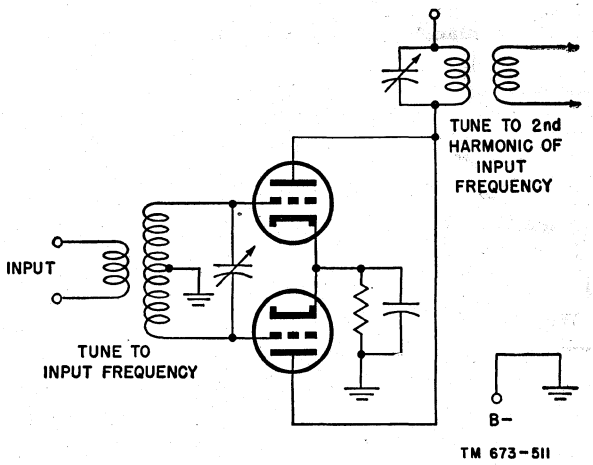


Figure 153. Push-push doubler circuit.

Section II. SOME OSCILLATORS USING CONVENTIONAL TUBES

96. Representative Feedback Oscillator Circuits

The usual radio-frequency signal source at frequencies below the microwave range is a feedback oscillator of the type described in paragraph 93. This section takes up a number of representative feedback oscillator circuits and discusses some of the factors that prevent their use in the microwave region. It includes also a discussion of the Barkhausen oscillator, a type of oscillator seldom used, but mentioned here because the discussion of its operation makes a good introduction to the subject of velocity-modulated tubes, which are taken up in paragraphs 101–105.

a. Basic Circuit of Feedback Amplifier. The basic loop configuration shown in figure 144 is illustrated again in A of figure 154, but the circuit aspect, rather than the feedback aspect of the basic circuit is emphasized in B. Here, the *useful output* connection has been omitted, and the feedback path is shown as a load network fed by the amplifier. The output in B is the feedback signal, which, when the loop is closed, supplies the amplifier input. Suppose, now, that the input to the amplifier is a signal, E, and the output

from the load network is a signal, E'. If there is a frequency for which E' and E are precisely in phase, with E' larger than E, the system will oscillate at that frequency when the loop is closed. It is necessary that E' be larger than E; otherwise the amplitude of oscillation will not increase. When the system is in stable operation, of course, the amplifier is saturated partially so that its gain is reduced. For signals at the normal oscillation level, E' and E are equal (they must be equal when the loop is closed, for then E *is* E'), but for low-level inputs with the loop open, E' is greater than E. The basic feedback amplifier, therefore, consists of an amplifier driving a load network, with the output of the load network serving as the input to the amplifier. The *useful output* normally is taken directly from the amplifier output. Usually the amplifier consists of a single stage, either single-ended or push-pull. The requirement on the load network is that *at the frequency of oscillation*, the output of the load network must be in phase with the input to the amplifier and (for low-level signals) must be slightly greater in amplitude.

b. Load Network Characteristics. A single tube can be replaced either by a voltage generator or by a current generator, both of which are discussed in appendix II. In figure 155, E_{gc} is the input voltage applied between the grid and the cathode of the tube and μ, R_p, and G_m are the tube constants. The two circuits, A and B, are equivalent, but it is probably desirable to concentrate on the voltage generator circuit, since it is a more familiar one than the current generator. In this circuit, the load network is fed from a resistive source having an emf (electromotive force) proportional to $-E_{gc}$. The problem of gain may be neglected for the moment, since it is usually easy to obtain the required gain of more than 1. The load network then is required to develop an output in phase with, and of slightly greater amplitude than E_{gc} when fed from a resistive source having an emf proportional to $-E_{gc}$. In other words, the load network, when fed through the generator impedance R_p, must develop a 180° phase shift at the frequency of oscillation. This requirement is illustrated in figure 156.

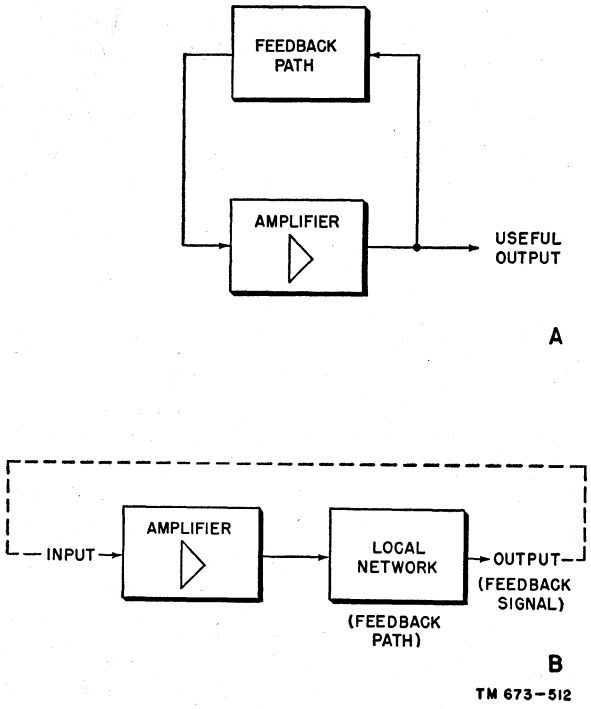


Figure 154. Basic circuit of feedback oscillator.

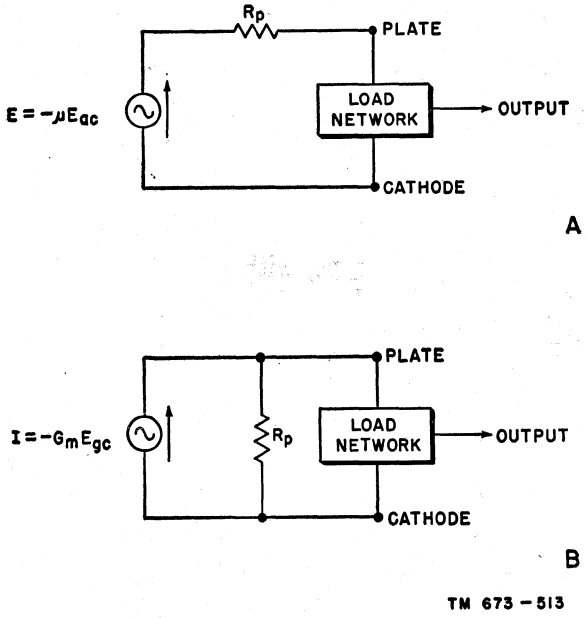


Figure 155. Equivalent amplifier circuits.

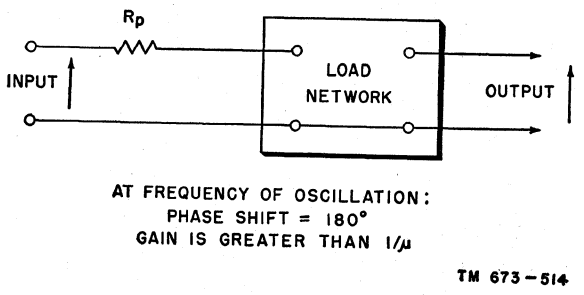


Figure 156. Load network characteristics.

c. Transformer-Coupled Oscillator. One of the simplest load networks that will supply the required phase shift of 180° is a transformer. The load network (A of fig. 157) (with the plate resistance R_p) is, then, nothing more than a simple air-core transformer with primary tuned so that it appears resistive and supplies precisely the desired phase shift. B shows the circuit of an oscillator of this type, with grid-leak bias. The parallel resistor (the grid leak) and capacitor act as an automatic bias circuit. During positive swings, the grid draws current, since it is more positive than the cathode, and charges the capacitor. During the remainder of the cycle, the capacitor discharges through the resistor. The time constant of the resistor-capacitor combination is much longer than a cycle, however, and the capacitor accumulates a charge in the direction shown on the drawing. On this account, if the amplitude of oscillation increases, the capacitor acquires a greater charge and applies additional negative grid bias, which reduces the gain of the system. By proper adjustment of the grid leak and grid capacitor as well as by proper choice of the voltage ratio in the feedback transformer, an operating point can be chosen at which the amplifier is reasonably linear so that the oscillator output has a good waveform.

d. R-C Oscillator. Another simple load network that provides the required phase shift is

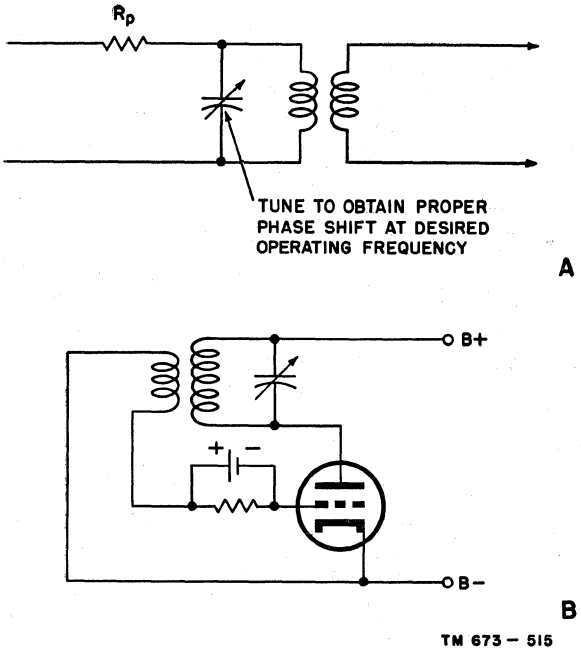


Figure 157. Transformer coupled oscillator circuit.

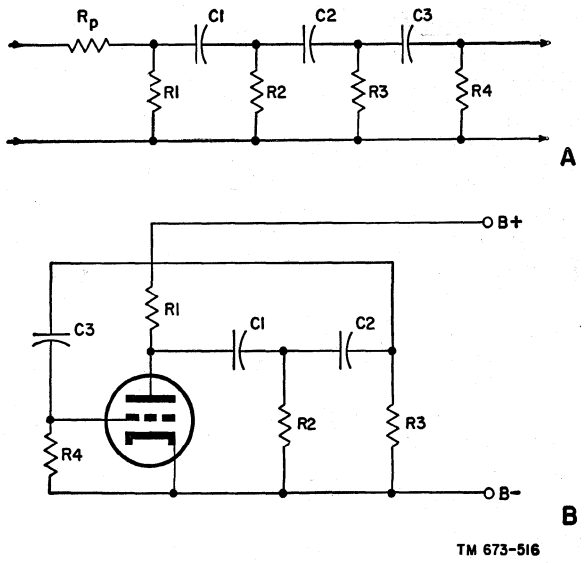


Figure 158. R-C oscillator circuit.

composed of a number of resistor-capacitor sections (A of fig. 158). The circuit of the oscillator, shown in B, often is used in the audio-frequency range.

e. Three-Reactance Network. This is the most frequently used load network (fig. 159) particularly at medium and high frequencies. In A, it is shown with a single capacitor, and in B it is shown with two capacitors. The theory of these networks is simple. If, in A, the reactance of C is greater than the reactance of L_1, the branch C-L_1 will look capacitive. The current through the branch then will lead the voltage at X by 90°. The voltage across L_1, however, always leads by 90° the current through L_1, and it is therefore 180° out of phase with the voltage at X. The output of the load network is, therefore, 180° out of phase with the voltage at X. If, then, L_2 is chosen so that its inductive reactance is equal (but of opposite sign) to the capacitive reactance of branch C-L_1, the impedance at X will be resistive and the voltage at X will be in phase with the voltage at the input. Similarly, in B, if the impedance of L is greater than that of C_1, the branch will look inductive. The current in branch L-C_1 will, therefore, lag the voltage at X by 90° and, of course, the voltage across C_1 lags the current by 90°. The result is a 180° phase shift between the voltage at X and the output voltage. If C_2 is chosen properly, the impedance at X will be resistive at the oscillator frequency, and the network will provide the required 180° phase

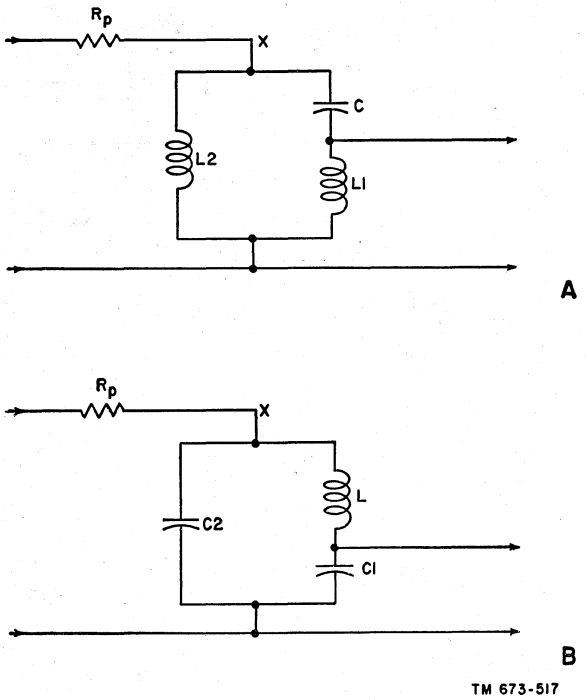


Figure 159. Three-reactance networks.

shift between input and output voltages. The basic circuits of three-reactance oscillators are shown in figure 160, with one capacitor shown in A and two capacitors in B. The one capacitor circuit uses series feed at both plate and grid (the circuit element is in series with the d-c supply to the tube element), but the two-capacitor circuit is forced to use shunt feed for both plate and grid (the circuit element is in parallel with the d-c supply to the tube element) and requires a d-c blocking capacitor in series with the plate-grid inductor as well as an r-f choke for the plate feed and a resistor for the grid feed. A number of three-reactance oscillators are described below.

f. Tuned-Plate, Tuned-Grid Oscillator. A tuned-plate, tuned-grid oscillator, shown in figure 161, is a three-reactance oscillator with one capacitor. The two tanks, or tuned circuits, are adjusted to look inductive at the oscillator frequency and to function as the inductors in a three-reactance network. The third reactance is the capacitor between the plate and the grid. This capacitor is required when a pentode is used but, when a triode is used, the plate-grid capacitance of the tube is usually sufficient to complete the network and the capacitor may be omitted.

g. Hartley Oscillator. The Hartley oscillator

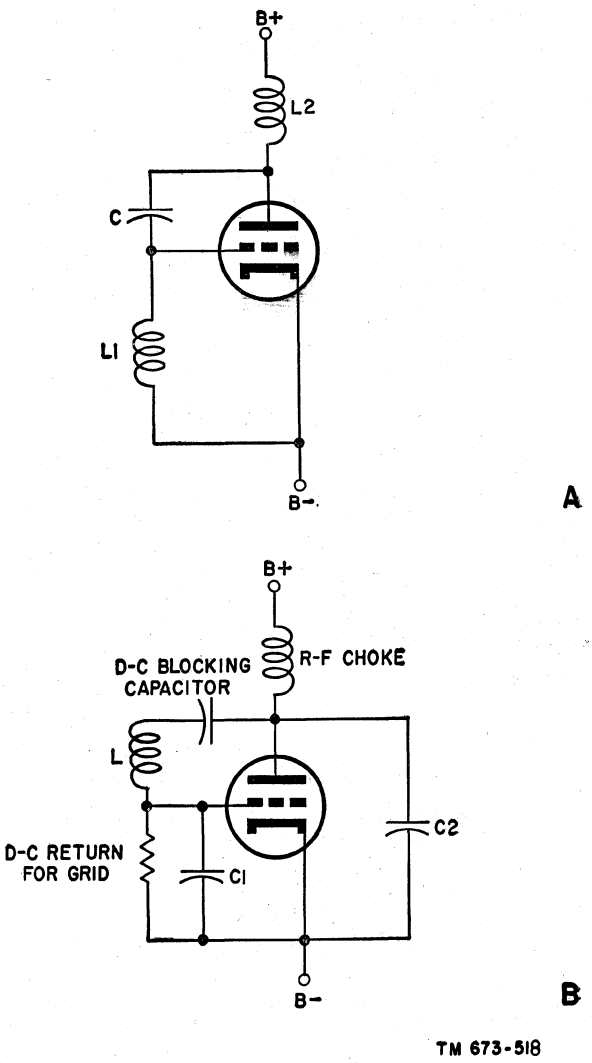


Figure 160. Three-reactance oscillator circuits.

(fig. 162) is another three-reactance oscillator with one capacitor. The two inductors, however, are wound as a single coil so that shunt feed must be used for the plate. This requires the r-f choke and the d-c blocking capacitor shown in the drawing. The Hartley oscillator is simpler than the tuned-plate, tuned-grid oscillator because it has only one tuning adjustment.

h. Colpitts Oscillator. The Colpitts oscillator (fig. 163) is a two-capacitor version of the three-reactance oscillator. In practice, the two tuning capacitors are ganged so that only a single control is needed. Since one side of each of the variable capacitors is grounded, the Colpitts oscillator is a little easier to construct, from a mechanical standpoint, than the Hartley oscillator.

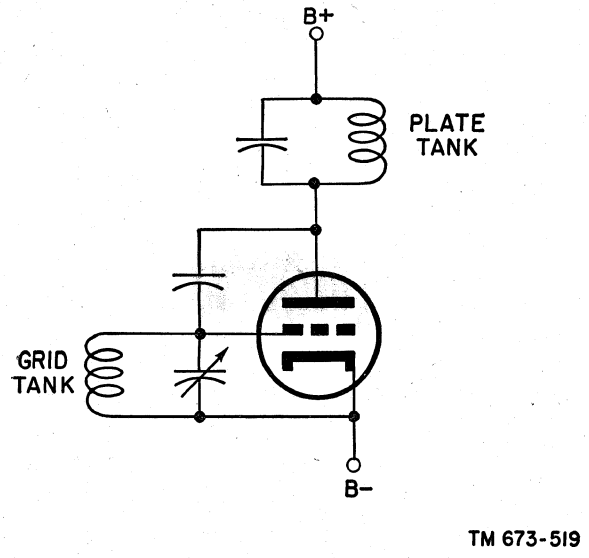


Figure 161. Tuned-plate, tuned-grid oscillator circuit.

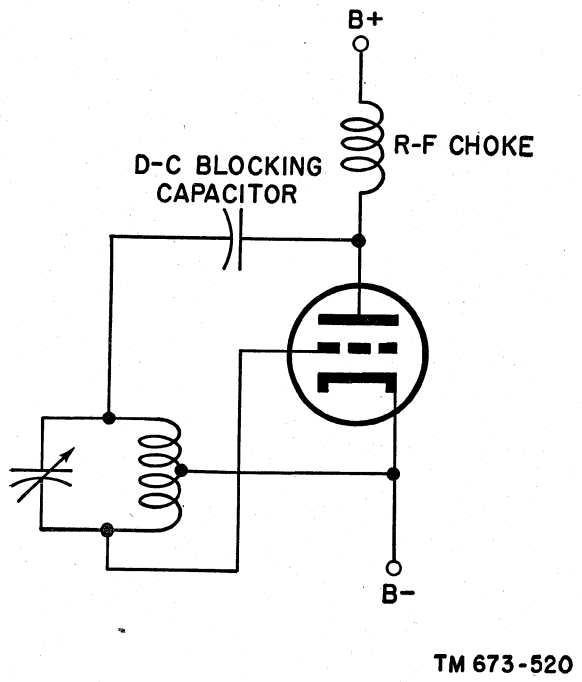


Figure 162. Hartley oscillator circuit.

i. Crystal Oscillator. A crystal oscillator is one that uses a quartz crystal as a frequency-controlling element. (This should not be confused with the crystal mixers and crystal rectifiers used at microwave frequencies.) The behavior of such a crystal is described briefly in appendix II. The equivalent circuit of the crystal is shown in A of figure 164, and the oscillator circuit is shown in

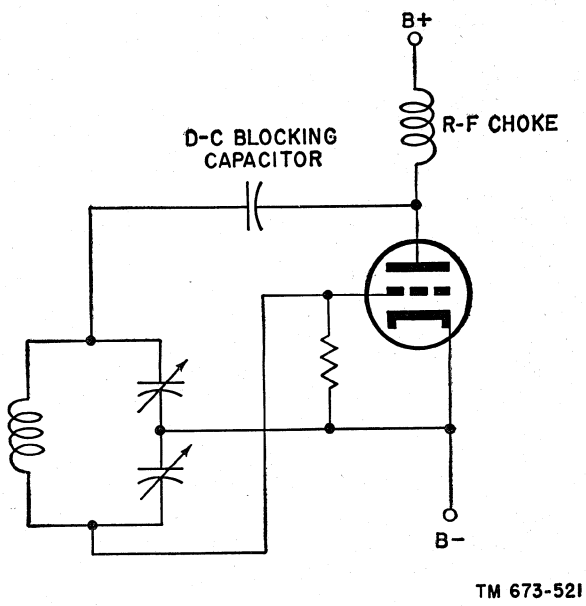


Figure 163. Colpitts oscillator circuit.

B. This is a three-reactance oscillator of the one-capacitor type and is electrically equivalent to the tuned-plate, tuned-grid oscillator described in *f*, above, and shown in figure 161. The required capacitance between the plate and the grid is supplied by the tube capacitance. The crystal, shunted by the r-f choke which provides a d-c grid connection, looks inductive at the operating frequency, and the plate tank is tuned to look inductive as well. The precision with which the quartz crystal controls the oscillator frequency is indicated in A, of figure 164. At the resonant frequency of the crystal, the *L-R-C* (inductance-resistance-capacitance) branch of the equivalent circuit is resistive. Below the resonant frequency, this branch looks capacitive. In either case, the equivalent circuit of the crystal, the *L-R-C* branch in parallel with *C′*, looks capacitive. Above the resonant frequency of the crystal, the *L-R-C* branch looks like an inductance. The crystal, however, can appear inductive only if the inductive current through the *L-R-C* branch is *greater* than the capacitive current through *C′*. Because of the high *Q* of the crystal, the inductive reactance of the *L-R-C* branch increases very rapidly with frequency (above the resonant frequency of the crystal) and there is only a small frequency range in which oscillation is possible. The *Q* of a crystal is something like 10,000. This is much higher than the *Q* of an ordinary tuned circuit.

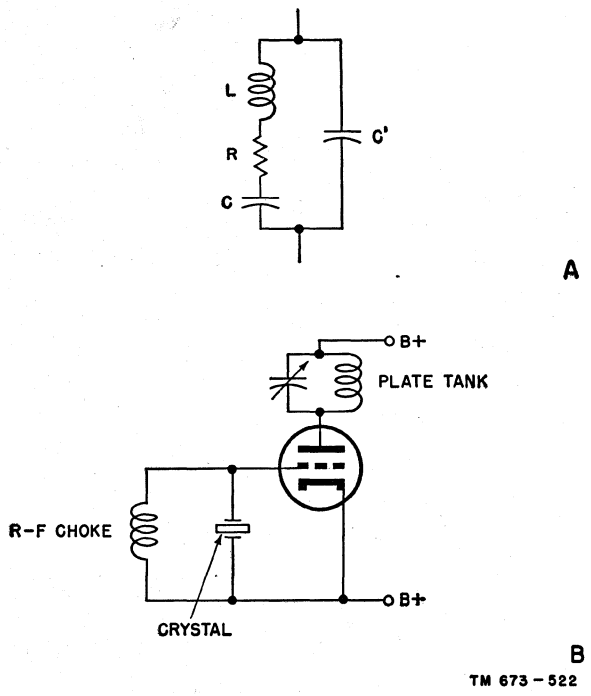


Figure 164. Crystal oscillator circuit.

97. Tube Capacitance and Lead Inductance

The feedback oscillator discussed in the previous paragraph appears to be a simple and uninvolved circuit. The situation is complicated, however, by the presence of capacitance between the elements of the vacuum tube and, to a lesser degree, by the inductance in the leads to the tube elements.

a. Transformer-Coupled and R-C Oscillators. Neither the transformer-coupled oscillator nor the *R-C* oscillator can be operated successfully except at relatively low frequencies. In these oscillators, as in all feedback oscillators, signals are transferred from the plate to the grid through the load network. Some additional signal also is transferred from the plate to the grid through the plate-grid capacitance of the tube, and this path does not provide the required 180° phase shift in either transformer-coupled or *R-C* oscillator circuits. In figure 165, A shows the circuit of a transformer-coupled oscillator, making use of the equivalent generator concept, with the plate-grid capacitance presented as though it were an external capacitor. B is the circuit of an *R-C* oscillator in the same way. These circuits will oscillate only when the effect of the plate-grid capacitance is very small. Since the reactance of the plate-grid capacitance decreases when the frequency is raised, high-frequency operation of either of these circuits is difficult or impossible because the bulk of the grid signal is supplied through the plate-grid capacitance rather than through the phase-shifting load network, and the resulting grid signal is not of the proper phase to sustain oscillation.

b. Three-Reactance Oscillators. Crystal oscillators cannot be used much above 10 mc per second because of the extremely thin crystals required for high-frequency operation. (Crystal-controlled transmitters are operated at high frequencies, but these involve oscillators followed by harmonic generators.) Other types of three-reactance oscillators can, however, be operated at frequencies close to the microwave region. The three-reactance oscillator with one capacitor is attractive for high-frequency operation, since the plate-grid capacitance becomes part of the load network and the other two tube capacitances may be tuned out by external inductors. A, of figure 166, shows the basic circuit of the oscillator, and B represents the circuit with the interelectrode capacitances added. Direct-current connections are omitted in each case. The conventional Hartley oscillator, using tubes having standard bases, shown in figure 167 with interelectrode capacitances added, operates well at moderately high frequencies. However, the possibility of operating such a Hartley oscilla-

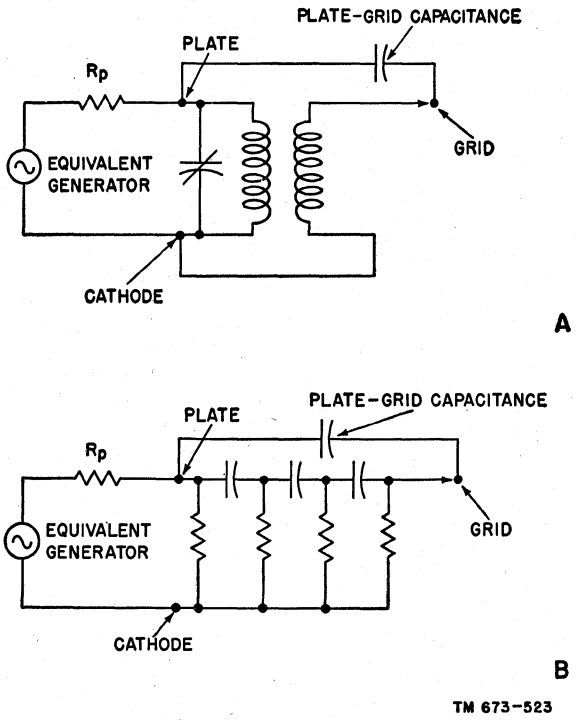


Figure 165. Oscillator circuits with plate-grid capacitance,

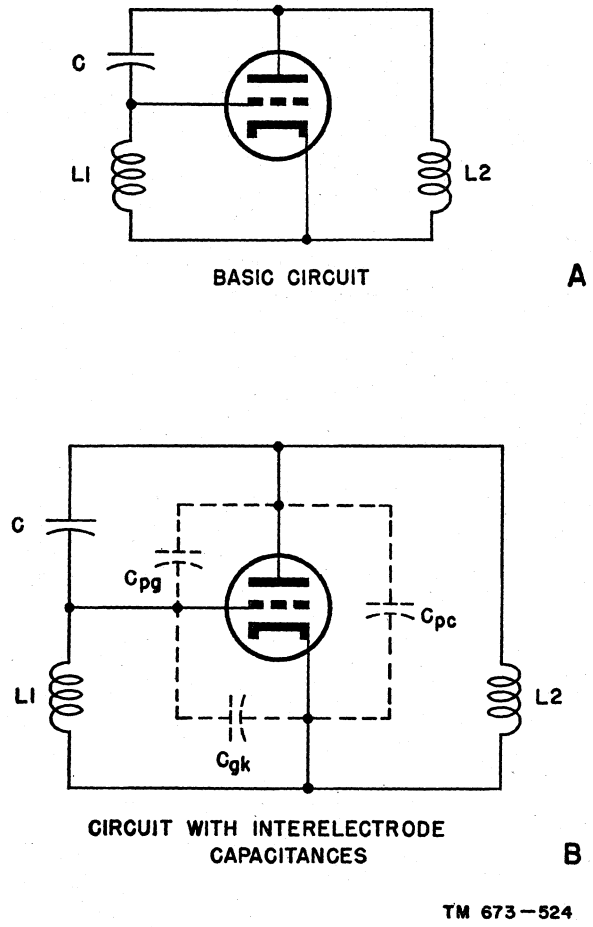


Figure 166. Effect of all tube capacitances.

tor at very-high frequencies is limited, because the external inductors are in series with the inductances of the tube leads. It is, therefore, impossible to obtain *little* enough inductance for vhf operation using ordinary tubes. This difficulty was solved in the middle 1930's by the development of the acorn tube. The early acorn tube series consisted of a triode, type 955, shown in A of figure 168; a sharp cut-off pentode, type 954; and a remote cut-off pentode, type 956, shown in B. The two pentodes are identical in external appearance, and all three tubes use indirectly heated cathodes operating at 6.3 volts. Since that time, other acorn tubes have been developed for battery operation. Interelectrode capacitances are minimized by small elements, and lead inductance is minimized by short straight leads. By virtue of this construction, the acorn triode can be used in

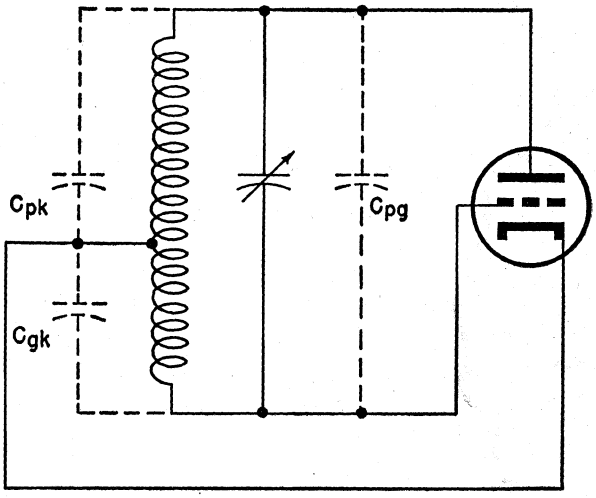


Figure 167. Hartley oscillator with interelectrode capacitances.

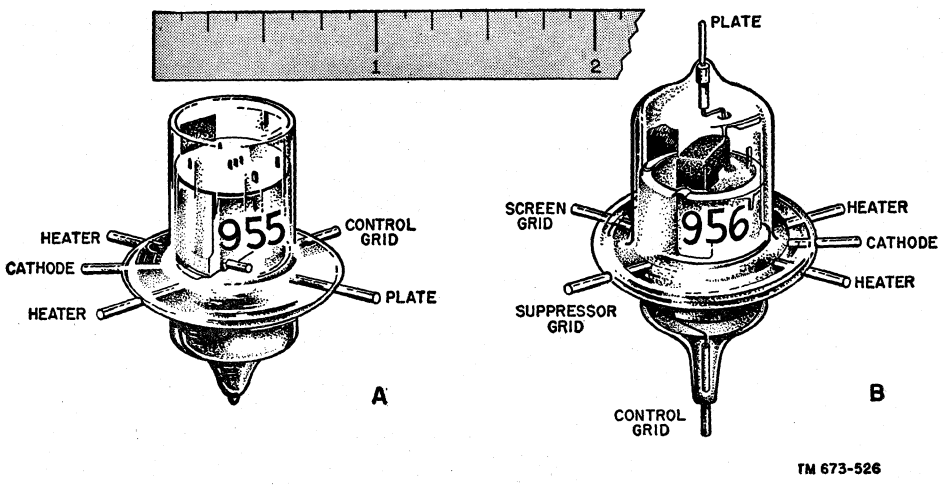


Figure 168. Types of acorn tubes.

a Hartley oscillator circuit at frequencies as high as 600 megacycles per second.

c. High-Power-Three-Reactance Oscillators. The miniature construction of the acorn tube limits the interelectrode capacitors, but unfortunately limits the power-handling capabilities as well. A large tube, having greater allowable dissipation than the acorn, and developing a correspondingly greater power output, cannot be operated at such high frequencies. A of figure 169, shows the basic three-reactance oscillator with one capacitor. It is clear from the circuit (discussed in par. 96*e*) that the required phase shift in the branch C-L_1 can be obtained only if the impedance of the branch is capacitive—that is, if the impedance of C is greater than that of L_1. Since the plate-grid capacitance is always present, this sets an upper limit on the frequency of the one-capacitor type of oscillator unless the circuit is modified by the addition of a shunt inductor that can be used to raise the plate-grid impedance. A simpler solution is the use of the two-capacitor circuit shown in B. It is convenient to operate the grid instead of the cathode at ground (A of fig. 170). The reason for this is illustrated in the push-pull circuit in B. The midpoint of the plate-to-plate inductor is essentially at ground potential, from the r-f standpoint, and the inductor thus is connected from plate to ground. This is, of course,

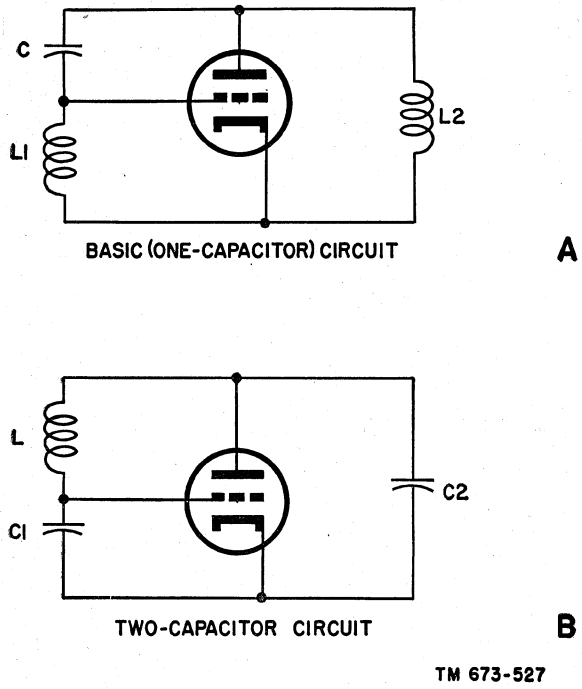


Figure 169. Three-reactance oscillator circuits.

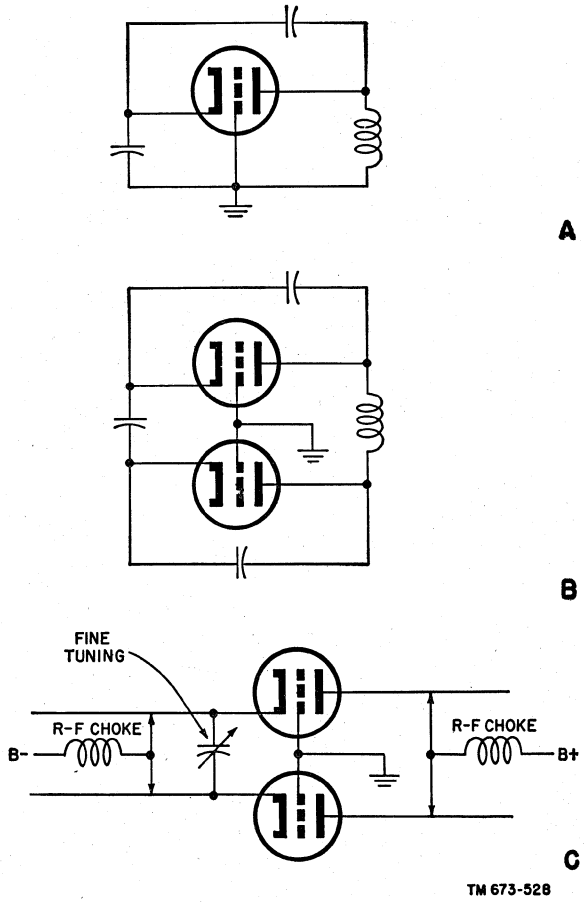


Figure 170. High-power oscillator circuit.

the same as a plate-grid connection if the grid is grounded. In practice, short lengths of transmission line may be used for the plate-to-plate inductor and the cathode-to-cathode capacitor, the lengths being chosen to obtain the proper impedance. Interelectrode capacitance in each tube supplies the plate-cathode capacitance. Fine tuning is accomplished by means of a low-capacitance trimmer across the cathode transmission line. An oscillator of this type, operated with a plate supply of 10,000 volts, is able to deliver up to 50 kilowatts in pulse service (during pulses) at frequencies of 200 mc per second. The tubes employed are type 8011, commonly called the micropup (fig. 171), especially developed for this application. A number of other conventional tubes are able to operate at frequencies higher than this. High-power oscillators similar to the one described immediately above can be operated easily at frequencies up to 600 mc per second, and acorn tubes can be operated above 1,000 mc per second.

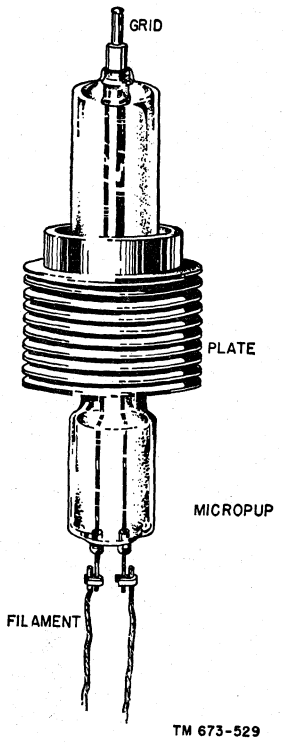


Figure 171. High-power high-frequency triode.

98. Transit Time

This discussion has shown that interelectrode capacitance and lead inductance are likely to be troublesome when an attempt is made to operate a conventional oscillator at high frequencies. However, with proper circuit design and the use of special tubes, which are conventional vacuum tubes so far as operating principles are concerned, the basic three-reactance, feedback oscillator can be made to deliver useful power at frequencies above 1,000 mc per second. The limiting frequency for a given tube, above which operation is not feasible, usually is set by the *transit time*—the time it takes an electron to travel from the cathode to the plate.

a. Effect of Transit Time. Because of transit time, the grid of a vacuum tube operated in the conventional way draws a resistive current. Thus, the effect of transit time on the external circuit is identical with the effect of a resistor connected between the grid and the cathode of the tube. Referring to figure 172, assume that the plate-cathode voltage and the grid-cathode voltage are constant. Assume, further, that the grid is sufficiently negative in relation to the cathode that no grid current flows. The plate current reaches a constant value shortly after the plate-cathode and grid-cathode voltages are applied, and, after this constant value has been reached, the number of electrons approaching the grid is the same as the number of electrons moving away from it. Now suppose that the grid is suddenly made more positive, but not so positive that grid current flows after the steady-state plate current has been established. Almost immediately the current flow will increase—more electrons will be moving—in regions X and Y, close to the grid. However, a short period of time will elapse before the current flow in region Z will increase. This time delay results from the fact that the grid can establish increased current flow only in its immediate neighborhood. Therefore, the current flow at Z cannot increase until the arrival, in that region, of the electrons that leave the grid at the instant when the grid voltage is changed. After the new value of plate current has been established, the number of electrons approaching the grid is the same as the number receding from it. *Before* the new value of plate current has been established, while the current at Z still has its original value, more electrons are approaching the grid than are moving away from it. This is true because the current change in region Z, which is the last change to occur (the current changes earlier in regions X and Y because they are closer to the grid), increases the number of electrons moving *away* from the grid and makes this number equal to the number of electrons approaching the grid. *The excess of approaching electrons constitutes a grid current, whether or not the approaching electrons ever*

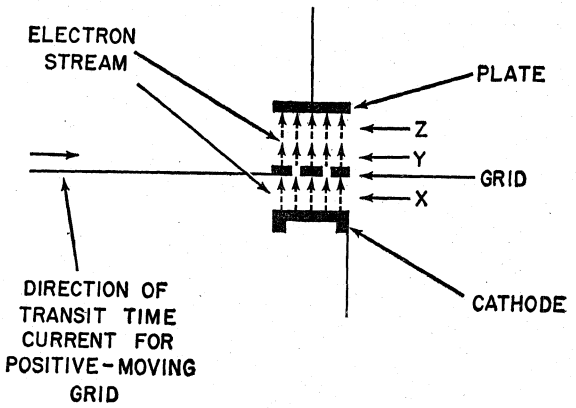


Figure 172. Effect of electron transit time on vacuum tube operation.

reach the grid, because these electrons repel electrons at the grid and cause an actual current of electrons leaving the grid by way of the grid lead. An increase in the grid voltage is, therefore, accompanied by a momentary flow of current *to* the grid (or a flow of electrons from the grid) in exactly the same way that an increase in the voltage across a resistor and a capacitor in series causes a momentary current flow. The effect of transit time is, therefore, the same as the effect of a series-R-C circuit connected between the grid and the cathode. When an alternating voltage is applied to the grid, the phase and magnitude of the transit time current depend on the relation between the transit time and the duration of 1 cycle. At moderate frequencies, the effect of transit time current is to increase the apparent grid-cathode capacitance. At higher frequencies the resistive effect is the important one, and the equivalent input resistance of the grid decreases with increasing operating frequency, other things being equal.

b. Effect of Transit Time on Oscillator Performance. Suppose that the amplifier, which is the heart of a feedback oscillator, delivers a power output, P. Some of this power is dissipated in the load network that feeds the grid, some of it supplies the required grid-driving power, and the remainder appears as useful output. With increasing frequency, the input resistance of the grid is decreased, because of transit time effects, and the power required to maintain a given voltage level at the grid is increased. At some frequency, the required grid-driving power, added to the power dissipated in the load network will be greater than P, and sustained oscillations will not be possible. That is, a tube output of P watts will not provide the grid-driving power, nor the grid voltage level, required to develop an output of P watts. This is simply another way of saying that the loop gain is less than 1.

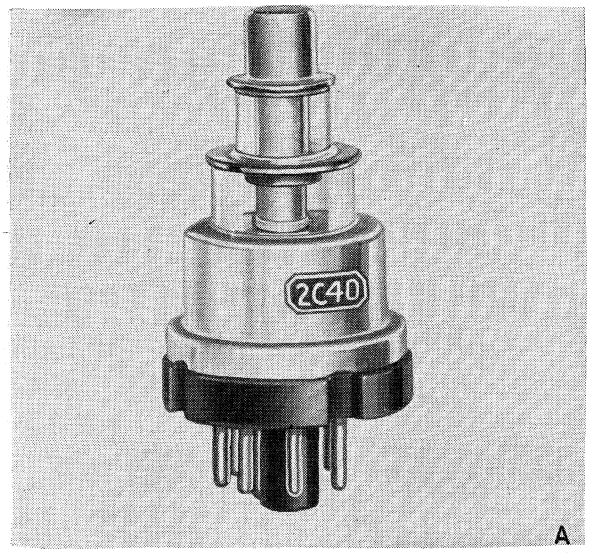
c. Methods of Decreasing Transit Time. There are two methods of decreasing transit time and thus increasing the maximum frequency of oscillation. One method is to reduce the spacing between the tube elements, particularly between the grid and the plate, so that the electrons have a shorter distance to travel. Reducing the spacing without making any other changes would increase the interelectrode capacitances and, therefore, it is necessary to make a corresponding reduction in the size of the tube elements in order to hold the capacitances within reasonable limits. This method, reduction in spacing and in size, is the one followed in the design of the acorn tubes and the lighthouse tubes. Unfortunately, reduction in the size of the tube elements effects a corresponding reduction in the power dissipation of the tube, so that use of this method tends to produce tubes having only low-power outputs. The second method, used in high-power transmitting tubes, is to increase the plate-cathode voltage so that electrons move from cathode to plate more rapidly. This method is followed in micropup and similar tubes. Plate voltages of 10,000 volts are common, and even higher voltages are not unusual.

99. Lighthouse Tubes

Most of the conventional high-power oscillator tubes used at frequencies near the microwave region are essentially similar to the micropup tube. They are triodes that rely on close spacing of the elements *and* on high anode voltage for reduction of transit time. In the low-power field, however, the acorn tubes have been superseded, for operation at the highest frequencies, by disk-seal or *lighthouse* tubes.

a. Construction. In most vacuum tubes, both the grid and the anode are in the form of open-ended cylinders that surround the cathode. The lighthouse tube (A of fig. 173) uses, instead, a series of plane or flat elements. The resemblance of this tube to a lighthouse is slight but nevertheless, during World War II, supplied its name. A cross section showing the construction of this tube is shown in B. The d-c cathode connection is through the base pins, and the r-f cathode connection is by way of the largest shell. The internal insulator between this shell and the cathode serves as the dielectric in a blocking capacitor, which keeps d-c voltage from the shell. The use of disk terminals reduces the lead inductance of the lighthouse tube to extremely low values. In addition, the internal construction allows the grid conductor to extend beyond the cathode and anode so that it acts as an r-f shield between them. By means of this design, the capacitance between the cathode and the anode is reduced to a small fraction of a micromicrofarad.

b. Lighthouse Oscillator. The lighthouse tube generally is used in a grounded-grid circuit. The anode-grid impedance is made inductive, the cathode-grid impedance is made capacitive (and adjustable), and the anode-cathode capacitance



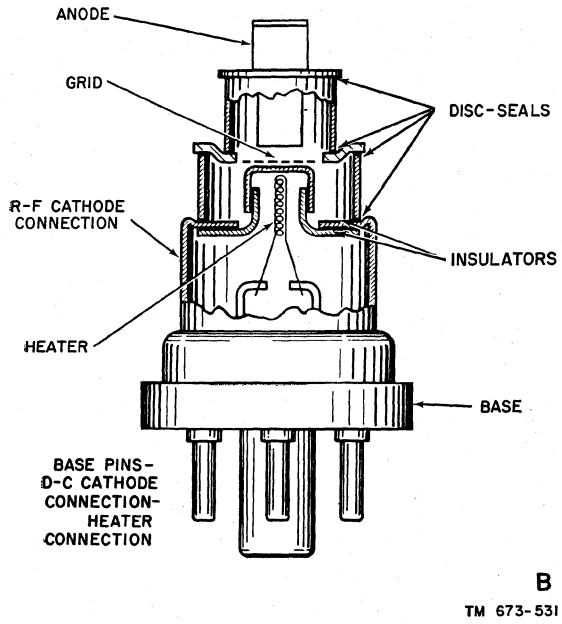


Figure 173. Lighthouse triode.

of the tube itself forms the third reactance of the feedback network. The tuning elements are ordinarily rigid coaxial lines (fig. 174). The outer conductor and the middle one form the cathode-grid line, and the middle and inner conductors form the anode-grid line. Both lines are tuned by means of adjustable shorting plungers. The output is taken from a coupling loop or a probe (not shown) which extends into the space between the anode and the grid conductors. The lighthouse tube frequently is used as an oscillator at

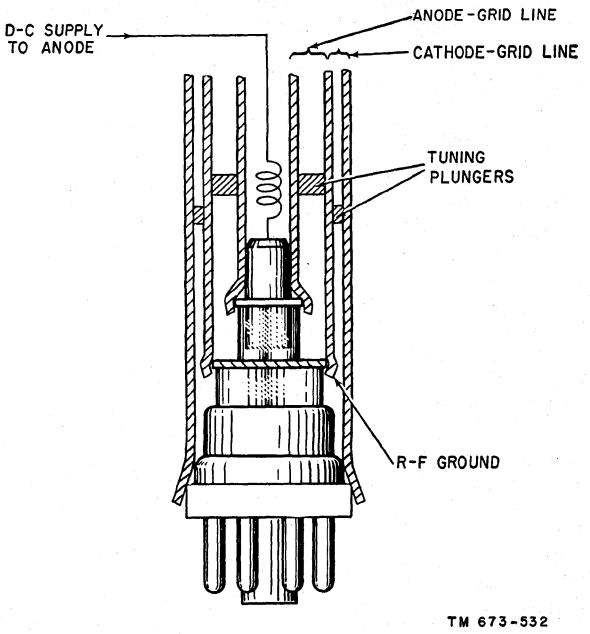


Figure 174. Lighthouse oscillator.

the lower microwave frequencies. It can be used also as an r-f amplifier, in a receiver.

100. Barkhausen Oscillator

Another type of oscillator, called the Barkhausen oscillator, also uses conventional vacuum tubes. This oscillator is not of any particular importance in military applications, but it is mentioned here because it has some similarity to the klystron oscillators used at microwave frequencies.

a. Basic Mechanism. Consider a conventional triode in which the anode is at the same voltage as the cathode, and the grid is at some positive potential. Electrons emitted by the cathode will, of course, be attracted to the positive grid. The grid, however, is constructed of very fine wires with large spaces between them, and most of the electrons will pass through the spaces. After passing through the grid structure, the electrons continue to be attracted by the grid, since it is more positive than the anode, and they reverse their direction and approach the grid again. The grid is, therefore, surrounded by a cloud of electrons, all of them oscillating through the plane of the grid, as shown in figure 175. As fast as electrons are captured by the grid and removed from the cloud, new electrons emitted by the cathode take their places. As explained in the discussion on transit time in paragraph 98, an electron approaching the grid is equivalent to a

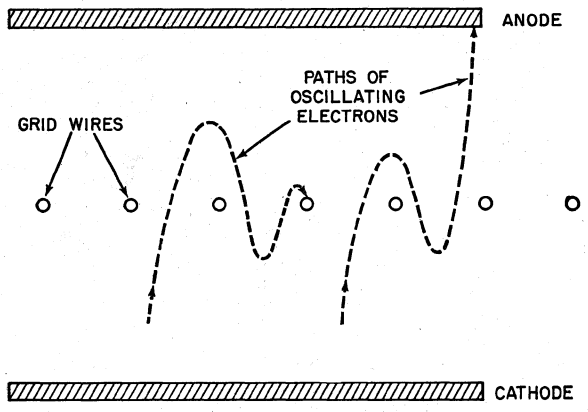


Figure 175. Barkhausen oscillations.

current *to* the grid, and an electron leaving the grid is equivalent to a current *from* the grid. There are, therefore, oscillatory currents of all possible phases in the grid line. The frequency of these, which is twice the frequency of the electron oscillations, is determined by the electrode voltages and by the spacing between the electrodes. The frequency of the oscillatory currents is twice that of the electron oscillations because the current direction is reversed *four* times during each electron oscillation cycle—once at each end of the electron path when the direction of the electron motion is reversed, and twice at the midpoint of the electron path when the electron passes through the grid structure.

b. Electron Sorting. Assume that an a-c voltage having the same frequency as the oscillatory currents is superimposed on the positive d-c grid voltage. These electrons that approach the grid during the positive half-cycle of the a-c grid voltage, and that leave the grid during the negative half-cycle, develop resistive currents and draw power from the generator that feeds the grid. That is, current associated with these electrons flows *to* the grid during positive half-cycles of the grid voltage and *away from* the grid during negative half-cycles. Electrons that oscillate 180° out of phase with these, however, deliver power *to* the grid, since the associated currents flow *from* the grid during positive half-cycles of the alternating grid voltage and *to* the grid during negative half-cycles. These electrons, therefore, deliver power *to* the generator that excites the grid. First, consider what happens to one of the electrons that delivers power to the grid. During the positive half-cycle of the grid voltage, this electron is moving away from the grid; during the negative half-cycles, it is moving toward the grid. Therefore, the electron is attracted strongly when it is leaving the grid and only weakly when it approaches the grid. The strong attraction as it moves away prevents it from escaping from the neighborhood of the grid, and the weak attraction as it approaches the grid prevents it from being captured immediately. Therefore, it continues to oscillate for a relatively long time. Now, consider the behavior of an electron that absorbs power from the generator that excites the grid. This electron is attracted strongly as it approaches the grid and only weakly as it leaves the grid. Therefore, it may be captured by the grid, or, alternatively, it may escape the influence of the grid entirely and travel either to the anode or the cathode. The result is that the electrons that *deliver* power to the grid continue to oscillate longer than those that *absorb* power from the grid.

c. Barkhausen Oscillator Circuit. The circuit of a typical Barkhausen oscillator is shown in figure 176. Any momentary inequality in the oscillatory currents in the grid line causes the resonant circuit to ring. This, in turn, delivers an a-c voltage to the grid, which sorts out the oscillating electrons according to the scheme described above, removing those electrons that absorb power. The oscillating electron cloud, therefore, continues to deliver energy to the resonant circuit, and stable oscillations result. The initial excitation of the resonant circuit results from tube noise—slight variation in the number of emitted electrons—which is present in all tubes.

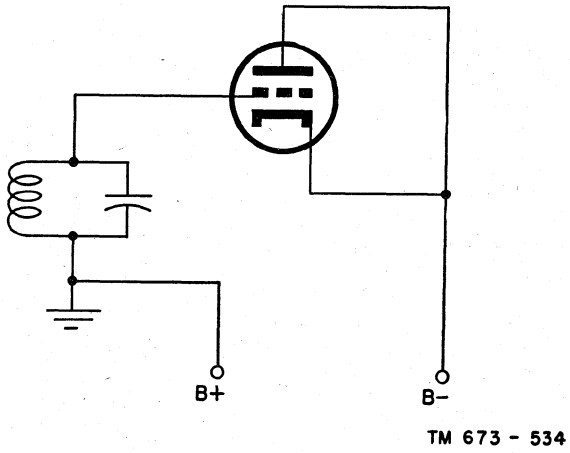


Figure 176. Barkhausen oscillator circuit.

Section III. KLYSTRON OSCILLATORS

101. Theory of Velocity Modulation

Velocity modulation, like almost everything else connected with microwaves, appears at first glance to be complicated, but actually is fairly easy to understand. Briefly, if a beam of electrons passes a fixed point, the nature of the beam may be derived from the number of electrons that pass the point during a given interval of time, and the velocity with which the electrons travel as they pass the point. In a conventional vacuum tube, operating in a conventional circuit, the electron beam is modulated by varying the number of electrons. In a velocity-modulated tube, the beam is modulated by varying the electron velocities.

a. Some Examples of Velocity Modulation. To clarify the concept of velocity modulation, men are used in the following example (fig. 177) as a substitute for electrons. Assume that, in a prisoner stockade, a military policeman passes prisoners through a gate on their way to a mess hall. As an example of an *unmodulated* beam, the MP passes two prisoners at a time through the gate at 1-second intervals, as in A, and all prisoners are marching at the same speed. At the gate, and at any other point between the gate and the mess hall, a steady stream of prisoners will pass

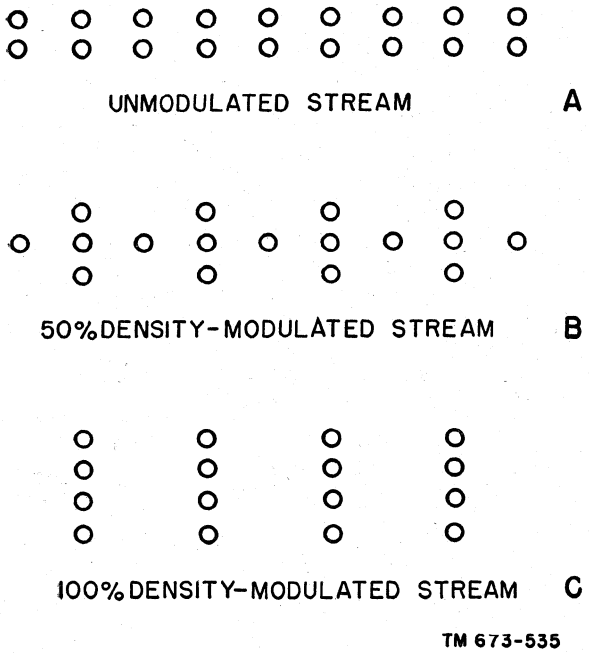


Figure 177. Examples of density modulation.

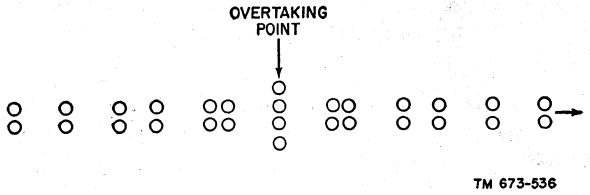


Figure 178. Example of velocity modulation.

at the rate of two per second. As an example of a *density-modulated* beam, in which the *number* of prisoners is varied, suppose that the MP continues to pass prisoners through the gate at 1-second intervals, but, at this time, he passes one prisoner, then three prisoners, one prisoner, three prisoners, and so on. In radio terms, the stream of prisoners now is density-modulated. An observer sees prisoners passing alternately at rates of one per second and three per second, as in B. The modulation level is, of course, 50 percent. That is, the maximum number per second is 50 percent greater than the number per second in the unmodulated stream, and the minimum number per second is 50 percent less than that in the unmodulated stream. A modulation level of 100 percent is achieved if the MP passes four prisoners through the gate at 2-second intervals, four prisoners at a time, as shown in C. The rate is then alternately zero and four per second. As an example of a *velocity-modulated* beam, suppose that the MP passes prisoners through in pairs at 1-second intervals, as in the unmodulated stream, but that the first pair of prisoners marches more slowly than usual, the next pair marches faster than usual, the next pair slower, the next pair faster, and so on. At the gate, the stream appears to be unmodulated but, as the prisoners march, the slow ones fall back and the faster ones move ahead in the stream. At some point, each pair of fast-marching prisoners will overtake the slow pair ahead and, *at this point*, the prisoners will pass an observer in groups of four, one group every 2 seconds (fig. 178). At some further point, each faster pair will overtake the *next* slower pair ahead of it, and an observer will see the same thing. Therefore, if the velocity-modulated stream is observed at the proper point, and there are an infinite number of proper points, it will appear to be density-modulated.

b. Applegate Diagram. The behavior of a velocity-modulated beam commonly is shown on

a graph called an *Applegate diagram* (fig. 180). In figure 179, A shows the distance traveled by an electron as a function of time. The electron passes a given point (which is the source of the velocity-modulated beam) at a time that is taken as $t=0$. It then travels with constant velocity, so that its distance from the point increases uniformly with time. B shows similar plots for a number of electrons that leave the given point at intervals after the time $t=0$. Suppose, now, that the beam is velocity-modulated. Some electrons will travel faster than others, and the plots for those electrons will be more nearly vertical than the plots for the slower ones as shown in figure 180. The points where *bunching* occurs—where the faster electrons overtake those traveling at normal velocity *and* where the slower electrons fall back to those traveling at normal velocity—are shown clearly in this diagram.

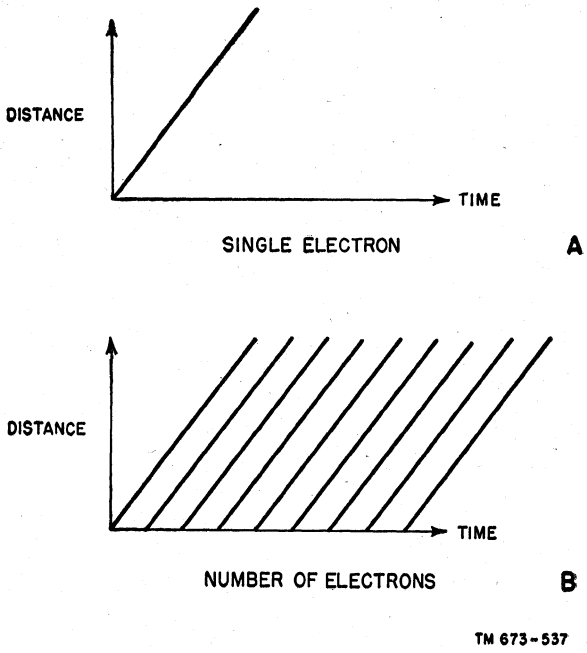


Figure 179. Electron time-distance plots.

c. Some Velocity-Modulation Calculations. Consider a beam of electrons that is velocity-modulated at a frequency of f cycles per second. Let V be the speed of the standard-velocity electrons, let $V(1+k)$ be the speed of the fastest electrons, and let $V(1-k)$ be the speed of the slowest electrons. Finally, let T be the duration of a cycle, so that $T=1/f$. Suppose that one of the fast electrons leaves the beam source at a given instant. The slow electron immediately ahead left the beam source 1 half-cycle earlier, and has already traveled a distance $V(1-k)T/2$. The preceding slow electron left the beam source 3 half-cycles earlier, and has traveled three times as far, and so on. The fast electron moves at a speed of $V(1+k)$, and the slow electrons move at a speed of $V(1-k)$. Therefore, the fast electron overtakes all of the slow electrons, and its over-

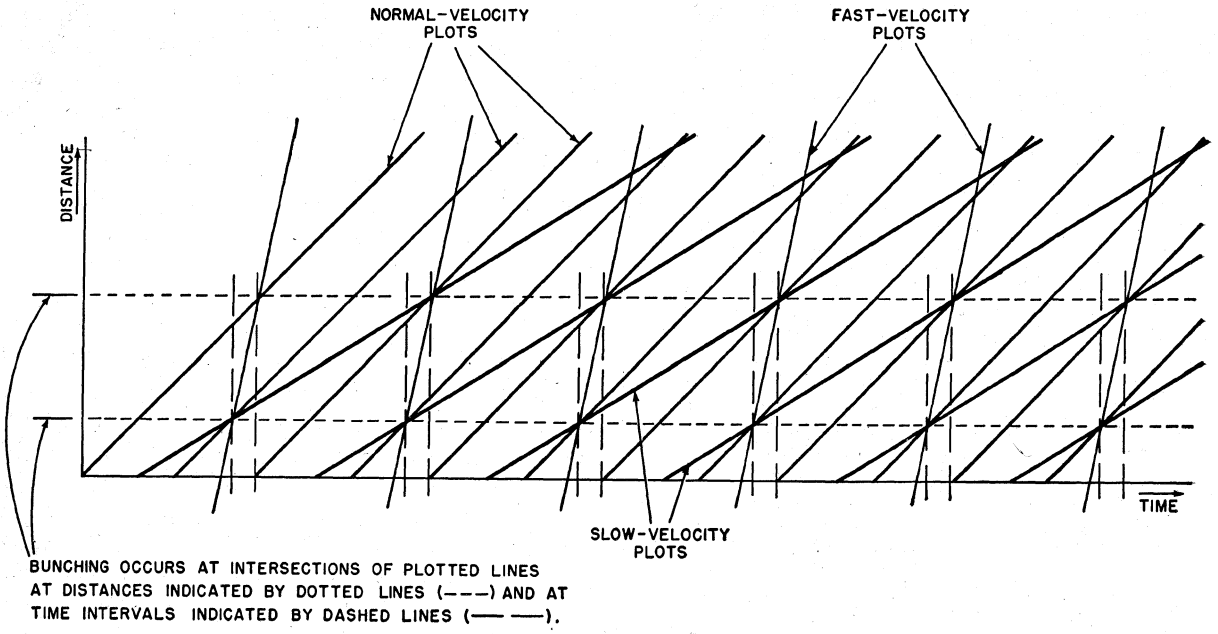

NORMAL-VELOCITY PLOTS
FAST-VELOCITY PLOTS
DISTANCE
TIME
SLOW-VELOCITY PLOTS
BUNCHING OCCURS AT INTERSECTIONS OF PLOTTED LINES AT DISTANCES INDICATED BY DOTTED LINES (– – –) AND AT TIME INTERVALS INDICATED BY DASHED LINES (— — —).
TM 673-538

Figure 180. Applegate diagram.

taking speed is $2kV$. The time required to overtake the first slow electron is $V(1-k)T/2$ divided by $2kV$, or $(1-k)T/4k$. The second slow electron is overtaken after a time interval three times as long as $(1-k)T/4k$, the third one, after a time interval five times as long, and so on. The distance traveled by the fast electron is $V(1+k)$ times the time interval, so that the first slow electron is overtaken after the fast electron travels a distance $V(1+k)(1-k)T/4k$ or $V(1-k^2)T/4k$. The second slow electron is overtaken after three times this distance, the third one after five times the distance, and so on. Therefore, bunching takes place at distances from the beam source that are *odd* multiples of $V(1-k^2)T/4k$. Since k is usually small, the quantity $(1-k^2)$ is nearly equal to 1, and the bunching distances may be taken as odd multiples of $VT/4k$.

102. Simple Klystron

A simple klystron consists of four parts: a *beam source*, a velocity-modulating unit called a *buncher*, a *drift tube* through which the velocity-modulated beam travels, and a unit called a *catcher*, which takes energy from the velocity-modulated beam.

a. Beam Source. The klystron beam source, shown in figure 181, consists of a cathode and its heater, a control grid, and an accelerating grid. The control grid, as is usual, controls the number of electrons in the beam. It does *not*, however, have any signal superimposed on the bias voltage. The accelerating grid speeds up electrons that are passed by the control grid. The velocity, V, is adjusted by varying the voltage on the accelerating grid.

b. Buncher. The buncher (fig. 182) is a resonant cavity, usually of modified cylindrical form, that has two bunching grids through which the electron beam passes. When microwave energy is fed to the buncher cavity (the input to the cavity is not shown in fig. 182), oscillations are set up, and an alternating voltage appears between the grids. When the second grid is more positive than the first grid, electrons passing through the buncher drift tube are speeded up. Conversely, when the second grid is more negative than the first grid, electrons passing through the buncher are slowed down. Therefore, the electron beam is velocity-modulated as it passes through the buncher (provided that the buncher cavity is fed with microwave energy). It is necessary, of course, to tune the cavity, and tuning paddles and plugs are provided. The cavity is made re-entrant in order to reduce the spacing between the two bunching grids. Thus, if an electron remains in the space between the grids for a whole cycle, it will be speeded up during part of the cycle and slowed down during the remainder. The spacing is reduced, therefore, and the electron velocity is made large, so that the time required for an electron to pass between the buncher grids is only a small fraction of a cycle.

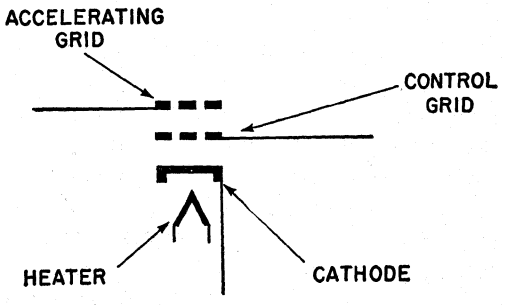


Figure 181. Klystron beam source.

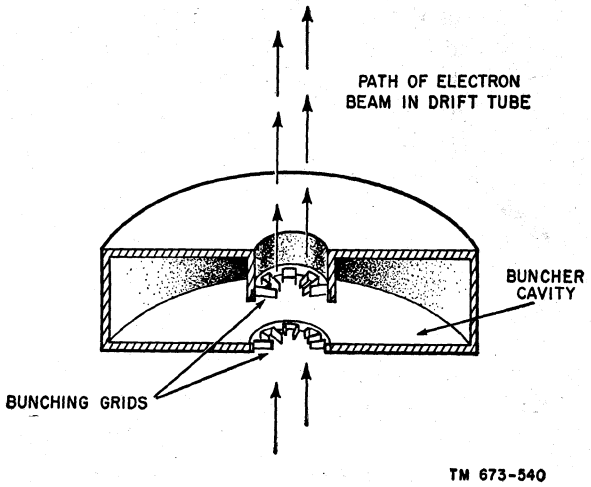


Figure 182. Klystron buncher.

c. Drift Tube and Catcher. The drift tube (fig. 183) is simply an evacuated tube through which the electron beam moves, and the catcher is a cavity with two grids like the buncher. In practice, the buncher and the catcher are almost identical.

d. Operation. In figure 183, all parts of a complete klystron are shown separately, including a collector electrode on which the electrons land after passing through the catcher grids. In practice, the cathode is made negative, and the

remainder of the tube, with the exception of the control grid, is operated at ground potential. The second catcher grid frequently serves as a collector, and the first buncher grid often serves as an accelerating electrode. Operation of the klystron is simple. An electron beam is passed through the buncher and is velocity-modulated by feeding energy to the buncher cavity. As the beam travels down the drift tube, the electrons are bunched; when the tube is adjusted properly, bunching occurs at the catcher grids. Electrons then pass through the catcher grid in bunches at the microwave frequency. It has been pointed out that an electron current is equivalent to any other current, and an alternating current between the grids of the catcher will excite the catcher cavity. Therefore, the tube acts as an amplifier, taking energy in at low levels at the buncher cavity and delivering an output from the catcher cavity. If some energy is fed back by means of a short length of coaxial cable from the catcher cavity to the buncher cavity, the klystron will oscillate when it is adjusted correctly.

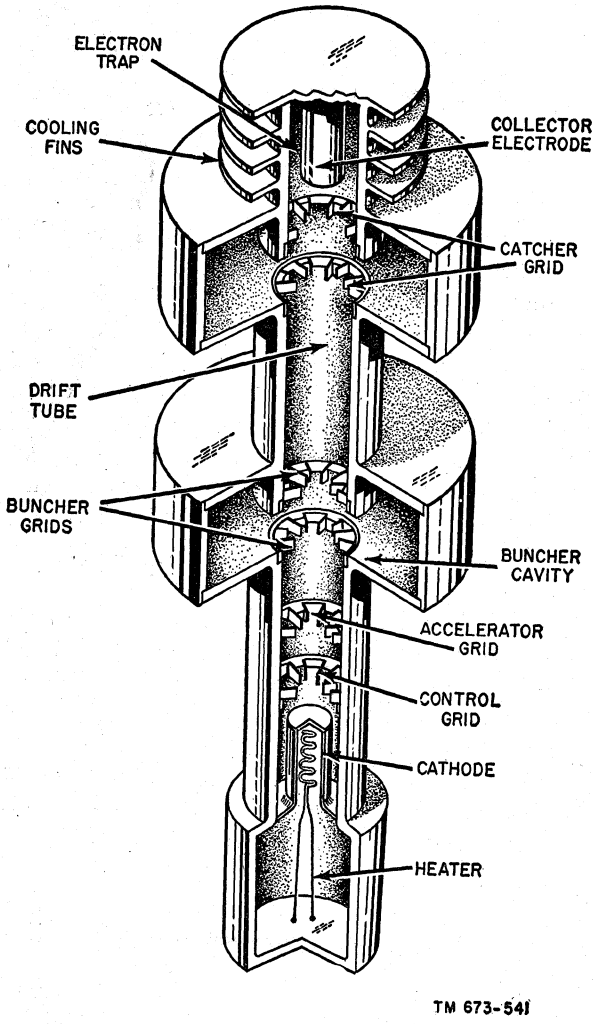


Figure 183. Klystron.

e. Operating Modes. It is necessary to tune both resonant cavities of the klystron before it will operate, either as an amplifier or as an oscillator. Besides this, however, it is necessary to insure that bunching takes place at the catcher grids. This is controlled by adjusting the negative voltage on the cathode, maintaining a fixed voltage difference between the cathode and the control grid, so that the accelerating voltage, between the cathode and the accelerator, brings the electrons to the proper velocity, V. As stated above, bunching occurs approximately at odd multiples of distance $VT/4k$, where the source of the modulated beam is the second buncher grid. Therefore, the distance between the buncher and the catcher must be an odd multiple of $VT/4k$ for the tube to amplify, or oscillate; otherwise it will neither amplify nor oscillate, since the electron beam delivers little or no energy to the catcher cavity. The klystron is said to be operating in the first mode (or in the fundamental mode) when the distance is approximately $VT/4k$, in the second mode when the distance is $3VT/4k$, in the third mode when the distance is $5VT/4k$, and so on. Bunching is less precise in the higher modes because of collisions between the electrons; fast electrons slow down on colliding, and slow electrons gain speed. Consequently, a klystron normally is operated in either the first or the second mode. It is possible to amplitude-modulate the klystron by varying the voltage on the control grid, and to frequency-modulate it by changing the accelerating voltage. In practice, however, klystrons of this sort seldom are used, since the reflex klystron, described in paragraph 103, is smaller and more efficient.

f. Mechanism of Oscillation. The klystron can function as an amplifier, as mentioned above, but it seldom is used, because of a fundamental difficulty. For successful amplification, the distance between the buncher and the catcher must be an odd multiple (more or less) of distance $VT/4k$. Unfortunately, k changes with the input level, and, with constant spacing, the klystron may be a good amplifier at one level and a poor one for signals at the same frequency but at another level. This characteristic, which makes the klystron of almost no value as an amplifier, gives it stability as an

oscillator. Consider the mechanism of oscillation: Some roughness in the electron beam excites the buncher cavity directly, or excites the catcher cavity and, through the coupling cable, the buncher cavity. In either case, the buncher cavity rings at its resonant frequency and velocity-modulates the electron stream. At a low signal level, k is small and the bunching distance is well beyond the catcher grids (since $VT/4k$ is larger than it should be). If k is large enough—if the shock excitation of the cavity has been sufficiently severe—the klystron will amplify, and the signal will be fed back from the catcher to the buncher. As the signal level is increased, k becomes larger, and the bunching distance becomes more nearly correct; therefore the amplification of the klystron increases. At some point, the bunching of the electrons takes place at the catcher grids, and the klystron operates at maximum gain. Further increases in k move the bunching distance away from the catcher grids again (still toward the buncher grids, by making $VT/4k$ smaller) and reduce the amplification. Just as in a conventional feedback amplifier, the level of oscillation increases (and the loop gain decreases) until a stable operating level is reached with a loop gain of 1. As an oscillator, the klystron operates well, seeking and maintaining a stable operating level. The klystron can be amplitude-modulated by changing the control grid voltage (which alters the number of electrons in the velocity-modulated beam). It also can be frequency-modulated by changing the acceleration voltage. Considerable interaction occurs between the two types of modulation, however, and neither type is completely successful because there is no easy method of varying the phase shift through the feedback loop (the coaxial cable from catcher to buncher). In practice, the simple klystron has been displaced almost completely by the reflex klystron, which is smaller and simpler.

g. Another Point of View. Before discussing the reflex klystron, it is desirable to look briefly at the simple klystron from another point of view that is slightly more complicated. Consider the resonant system, which consists of the two cavities (buncher cavity and catcher cavity) coupled by a coaxial cable. The discussion of the Barkhausen oscillator has shown that the acceleration of an electron absorbs energy from the resonant system, and that deceleration of an electron feeds energy to the system. The resonant system then absorbs energy from all electrons that pass the catcher grids when the *second* catcher grid is *less* negative than the first. Because the velocity-modulated electron beam is bunched by the time it reaches the catcher grids, almost all of the electrons arrive at times when they can feed energy to the resonant system, if the phasing between the cavities is correct for the spacing and for the operating potentials; only a few electrons arrive at times when they absorb energy. Velocity-modulating in the klystron is, therefore, equivalent to electron-sorting in the Barkhausen oscillator. On this basis, fast electrons, during acceleration, absorb energy from the buncher and feed it back to the catcher. Slow electrons, however, feed energy to *both* cavities. Successful operation of the klystron thus depends on the proper relation between cavity tuning, cavity spacing, phase shift through the coaxial cable, electron velocity, and modulation level. It will be seen in the discussion in paragraphs 103–105 that the reflex klystron is simpler, and this is its major merit.

103. Reflex Klystron

a. Description. A reflex klystron (figs. 184 and 185) may be thought of as a simple klystron that has been folded back on itself so that one cavity acts as both buncher and catcher. Alternatively, it may be considered as a type of Barkhausen oscillator with a cavity resonator for a tuned circuit. It consists of a cathode and heater, a resonator cavity with two grids, and a reflector. In the type 723A/B, which is the usual X-band reflex klystron, the cavity is tuned by physical distortion. The cavity, as well as the body of the tube, is operated at ground. The cathode is negative, so that the first cavity grid serves as an accelerator and a control grid. The second cavity grid, on the side away from the cathode, modulates the electron beam. The

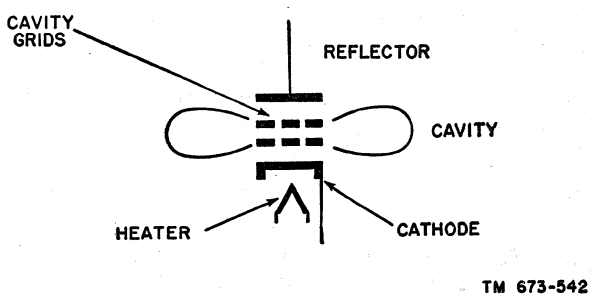


Figure 184. Reflex klystron, schematic view.

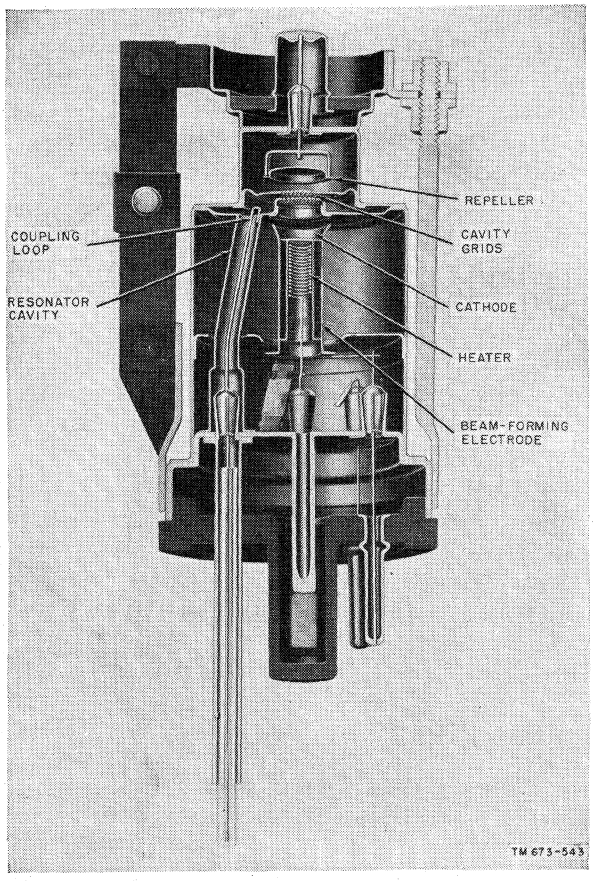


Figure 185. Reflex klystron, cutaway view.

reflector is operated at a voltage slightly more negative than the cathode, and it serves to turn electrons back to the cavity.

b. Operation. Operation of the reflex klystron depends on a proper relation between the cathode voltage, the reflector voltage, and the cavity tuning. Consider, for the moment, what happens to an electron that passes through the cavity grids at an instant when the second grid is at the same potential as the first and is going negative. The electron is repelled by the reflector and finally is brought to rest. After this, its velocity is reversed, and it moves back to the cavity grids, passing through and rejoining the cloud of electrons about the cathode. An electron that passes through the cavity grids slightly earlier, when the second grid is positive, approaches the reflector at a higher velocity and takes longer to stop. The result is that it arrives at the cavity grids, following reflection, after a longer trip. Similarly, an electron that passes through the cavity grids later than the first electron considered, when the second grid is negative, returns after a shorter trip. The result is a bunching process that is brought about by velocity modulation of the beam. Suppose, now, that a standard electron, which is not accelerated by the cavity grids, returns in ¾ cycle. An electron that passes the grids earlier and is accelerated returns in more than ¾ cycle. An electron that passes the grids later and is slowed down returns in less than ¾ cycle. As a result, if the system is tuned correctly, energy is delivered to the cavity, and the tube oscillates.

c. Operating Modes. As in the case of the simple klystron, it is possible for a reflex klystron to operate in a number of different modes. In the first mode, an unaccelerated electron returns to the cavity grids in ¾ cycle. In the second mode, it returns after 1¾ cycles. In the third mode, it takes 2¾ cycles to return, and so on. The frequency is controlled almost entirely by the reflector voltage, if the cathode voltage is held fixed, and tuning the cavity puts the system in and out of oscillation. As with the simple klystron, the operating level is fixed by the fact that gain is reduced when the level of oscillation rises above that level for which bunching occurs at the cavity grids. The system seeks and maintains the level of oscillation at which the energy input to the cavity just balances the sum of the losses and the energy output, to the electron beam *and* to the load.

104. Modulation Techniques for Reflex Klystrons

a. Change in Reflector Voltage. When the reflector is made *more negative* in relation to the cavity grids, the slowing force on the electrons is *increased.* Therefore, an electron passing through the cavity grids on its way from the cathode will return *sooner*, and the frequency of oscillation will be raised. Of course, there is a limit to the amount of frequency change that can be obtained by changing the reflector voltage. If the cavity is tuned initially to resonance, it will be slightly out of tune when the frequency is changed by variation of the reflector voltage. This will cause the output of the oscillator to drop. By the time the frequency has been shifted by about 20 or 30 mc per second, the power will have dropped to about one-half of the maximum value (obtained when the cavity is tuned). The frequency change is approximately 1 mc per second for a 1-volt shift in the reflector voltage. This variation in frequency is,

of course, distinct from the change that can be made by changing the reflector voltage *and* retuning the cavity. In general, by changing the reflector voltage and retuning the cavity, a reflex klystron can be *tuned* over a range of about plus or minus 5 percent from the center of its operating band. By changing the reflector voltage, it can be *frequency-modulated* over a range of about plus or minus .5 percent of its operating frequency. Note, however, that the frequency modulation is accompanied by some amplitude modulation when the frequency shift is accomplished by variation in the reflector voltage.

b. Change in Cathode Voltage. When the cathode is made *more negative* in relation to the cavity grids, electrons are accelerated *more* before reaching the grids. Therefore, they take *longer* to return to the cavity grids, and the frequency is *lowered*. At the same time, however, the number of electrons in the beam is increased so that the power output *rises* with *decreasing* frequency and falls with *increasing* frequency. The increase in the number of electrons in the beam, when the cathode is made more negative is similar to the effect of an increase in the plate current of an ordinary tube when the plate is made more positive.

c. Simultaneous Change in Cathode and Reflector Voltages. By changing both cathode and reflector voltages, although not by equal amounts, it is possible to achieve either amplitude or frequency modulation of a reflex klystron. If, for example, the cathode is made more negative, the power output increases and the frequency decreases. A slight increase in the reflector voltage will return the frequency to the initial value, or will prevent a frequency shift if the voltage changes are made simultaneously, without affecting the change in power level. Similarly, if the reflector voltage is altered to frequency-modulate the klystron, a slight change in the cathode voltage will compensate for cavity detuning and hold the output level constant.

d. Pulsed Operation. The reflex klystron will not operate if either the cathode or the reflector is at the same voltage as the cavity grids. It may, therefore, be pulse-modulated by applying negative voltage pulses to either the cathode or the reflector. It is usual, in order to avoid frequency variation during pulses, to supply voltage pulses to both the cathode and the reflector. Amplitude and frequency modulation of the pulsed output can be obtained by suitable modulation of the d-c pulses (negative) that are applied to these electrodes.

105. Applications of Reflex Klystrons

Reflex klystrons are used whenever microwave signals at low-power levels are required. For example, microwave signal generators use reflex klystrons with electrical tuning, by variation in the reflector voltage, and with mechanical tuning, by tuning of the cavity. Microwave receivers, which are almost always superheterodynes, use reflex klystrons as local oscillators. It is usual to take advantage of the possibility of electrical tuning by adding afc (automatic frequency-control) systems. These are described in chapter 6, which deals with microwave receivers. Another interesting application, which makes use of electrical tuning, is the *spectrum analyzer*. This is a device that sweeps a band of frequencies and displays the signal level (at each frequency) on an oscilloscope. The spectrum analyzer is nothing more than a narrow-band receiver the output of which is fed to the vertical deflection plates of an oscilloscope. An external sweep generator feeds a signal to the horizontal deflection plates of the oscilloscope and, at the same time, applies a saw-tooth voltage wave, superimposed on the fixed d-c voltage to the reflector of the klystron, which acts as a local oscillator for the receiver (fig. 186). The result is a picture on the oscilloscope screen in which frequency is shown on a horizontal scale (the horizontal deflecting voltage is also the tuning voltage for the receiver), and signal level at each frequency is shown on a vertical scale.

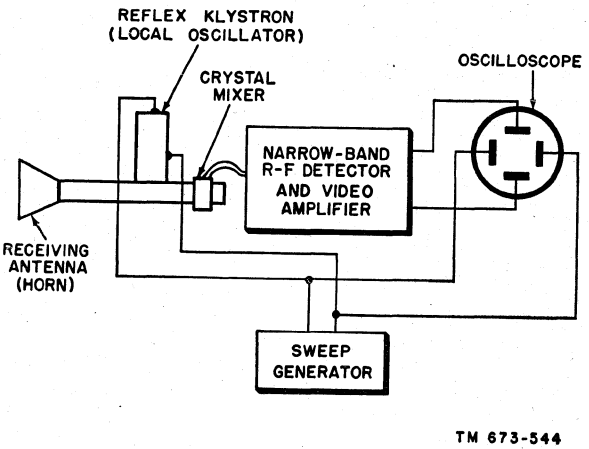


Figure 186. Spectrum analyzer.

Section IV. CAVITY MAGNETRONS

106. Basic Theory of Multicavity Operation

Before discussing cavity magnetrons, it is desirable to examine the behavior of some simpler systems which consist of several resonant cavities, such as the buncher and catcher cavities of a non-reflex klystron, and a single velocity-modulated electron beam.

a. Two-Cavity System. The two-cavity system (fig. 187) consists of two resonant cavities, both tuned to the same frequency, and an electron beam which passes through both sets of cavity grids. The figure does not show the source of the electron beam nor the evacuated envelope, since the system is of theoretical interest only. Suppose that the first cavity is excited at its resonant frequency by an external microwave power source connected to the coaxial input line. The electron beam will, of course, be velocity-modulated. If the system is adjusted properly, with correct electron velocity and modulation level and correct spacing between the cavities, the beam will be bunched at the grids of the second cavity. Under these conditions, the second cavity will resonate at the operating frequency, and a sizable signal will appear there. Also, the phase of the oscillation in the second cavity must be such that the bunched electrons are *slowed down*, for this is the only way that energy can be abstracted from the beam. Oscillation at any other phase either will be unaffected by the electron beam, if the bunched electrons pass through the cavity grids when both are at the same potential, or will deliver energy *from* the cavity *to* the beam, if the bunched electrons pass through the cavity grids at a time when they are accelerated. The electron beam, therefore, is able to carry energy from the first cavity to the second.

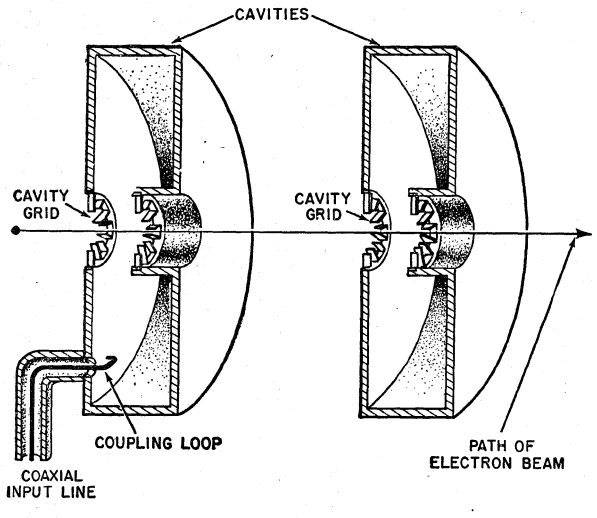


Figure 187. Two-cavity systems with electron-beam coupling.

b. Two-Cavity System With Underbunching. Consider now what happens if the second cavity is moved toward the first. The electron beam then is *underbunched* at the grids of the second cavity; that is, bunching is not yet complete. Since there is partial bunching, the flow of electrons is not smooth, and the second cavity is able to absorb energy from the beam. The phase of the oscillation in the second cavity automatically adjusts itself, by absorbing energy, so that the greatest possible number of electrons is slowed. The remaining electrons, which pass the cavity grids when the second grid is more positive than the first, are accelerated. The second cavity, therefore, slows down *more* than half of the electrons that pass through its grid, and accelerates *less* than half of them. The cavity, therefore, absorbs energy from the beam, reducing the average velocity of the electrons in the process, since it slows down more electrons than it speeds up, *and also* adds to the velocity-modulation of the beam.

c. Multicavity System. If a cavity system has more than two cavities, it is possible for *each* cavity, except the first one, to absorb energy from the beam and, at the same time, improve the velocity modulation of the beam for the next cavity. A small portion of the energy comes from the external source that feeds the first cavity. The greater part of the energy, however, is supplied by the electron beam source; each cavity reduces the average velocity of the electrons in the beam. Note that the system is extremely sensitive to adjustment, since the energy absorbed by each cavity depends on the relation between bunching distance and spacing between cavities.

d. Ring System. Suppose, now, that a number of resonant cavities, all tuned to the same frequency, are arranged in a ring (fig. 188). Assume, further, that a beam of electrons is made to travel in a closed loop that passes through the grids associated with each cavity. For the moment, the question of how the electron beam is made to take this path may be disregarded. Any roughness in the electron stream will cause one or more

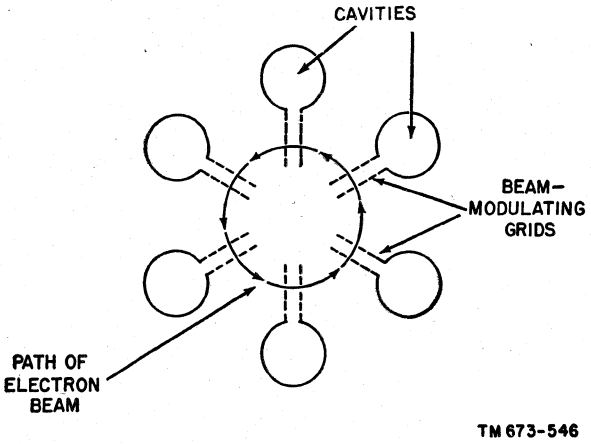


Figure 188. Ring system of resonant cavities.

of the cavities to ring. This will modulate the beam, and energy will be delivered to all of the cavities. The microwave energy that is delivered to the cavities comes, as before, from the original velocity of the electron beam, and transfer of energy to the cavities results in a reduction of the average electron velocity. *If matters can be arranged so that a continuous supply of fast electrons is available, while slow electrons are removed from the system, energy will be abstracted from the electron stream and will appear in all of the cavities.* (The energy in the cavities is, of course, r-f energy in the form of electric and magnetic fields.) The cavity magnetron operates in precisely this way.

107. Description

A cavity magnetron consists of a cylindrical anode in which a series of cavities are cut, a cathode, and a magnet. The magnet controls the paths of the electrons so that they travel in a more or less circular orbit.

a. Anode. Figure 189 shows one of the more common magnetron anodes, a *hole-and-slot* type. All of the types of magnetron anode are basically similar. A is a view along the axis of the cylinder that shows only the main outline of the anode. The reason for the name is obvious at once, since each of the cavities consists of a cylindrical hole joined to the central space by a small slot. Although six cavities are shown, magnetrons with other numbers of cavities are not uncommon. B is a sectional view of the anode in perspective showing the cooling fins on the outer surface of the anode and also showing the coupling loop and coaxial line, which takes energy from one of the cavities to the microwave transmission system,

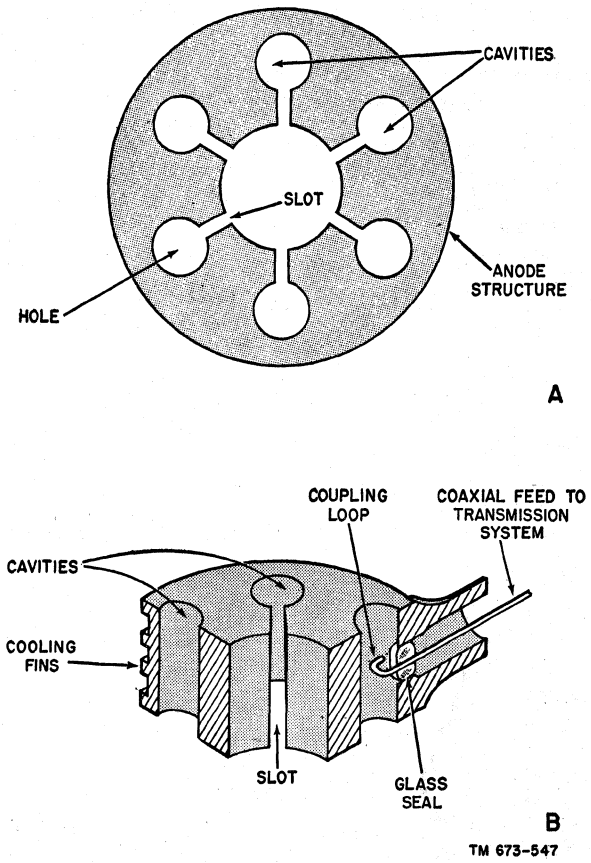


Figure 189. Hole and slot anode.

either waveguide or rigid coaxial line. There are no cavity grids. Instead, the electron beam is modulated by the field, which spreads out from the lips of the slot, as shown in figure 190. Other types of anodes are shown in figure 191. (A shows a *slot anode*, B, a *vane anode*, and C, a *rising sun anode*.) Choice of one anode type over another is based largely on the problems associated with construction of the magnetron, since all types behave in the same way.

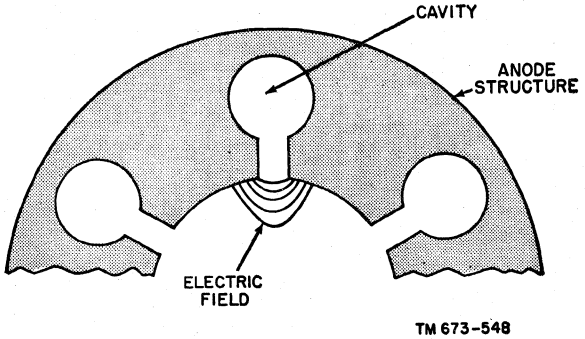


Figure 190. Field in neighborhood of slot.

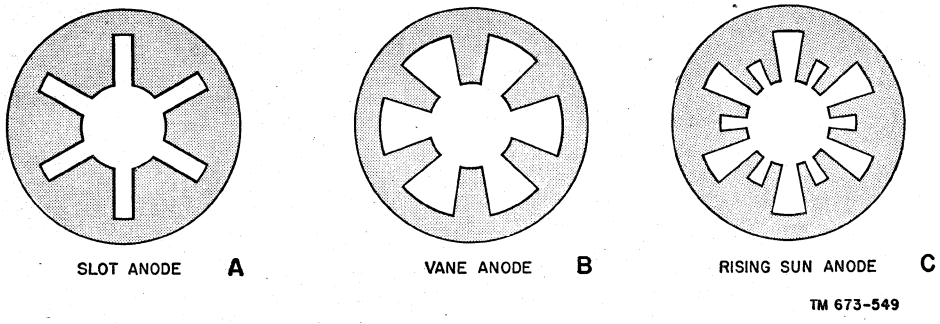


Figure 191. Types of anodes.

b. Cathode. The cathode of the cavity magnetron is placed along the axis of the anode cylinder. It is usually indirectly heated, and is connected electrically to one of the heater leads so that only two leads are required for heater and cathode (fig. 192). Most magnetrons are used in pulse service, and the cathode emission during pulses is in the neighborhood of 50 amperes, so a high-emission oxide almost always is used. It will be shown that many of the emitted electrons ultimately return to the cathode, and the energy that they give up contributes to the cathode heating, in the same way that the anode of a conventional vacuum tube is heated by the energy it absorbs from the electrons it receives. Because of this, the heater current is made high during warm-up of the equipment and is reduced during operation. In some high-power magnetrons, the heater even is turned off during operation, since the electrons captured by the cathode supply enough energy to maintain it at normal operating temperature. *The magnetron never must be operated without allowing sufficient time for the cathode to come to operating temperature.* If this precaution is not observed, the magnetron may be damaged.

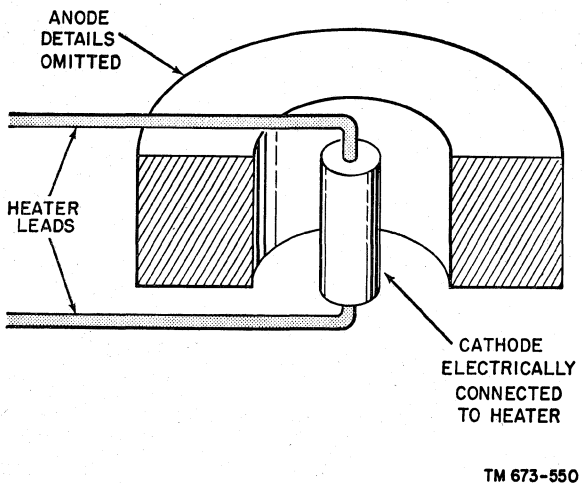


Figure 192. Cathode of magnetron.

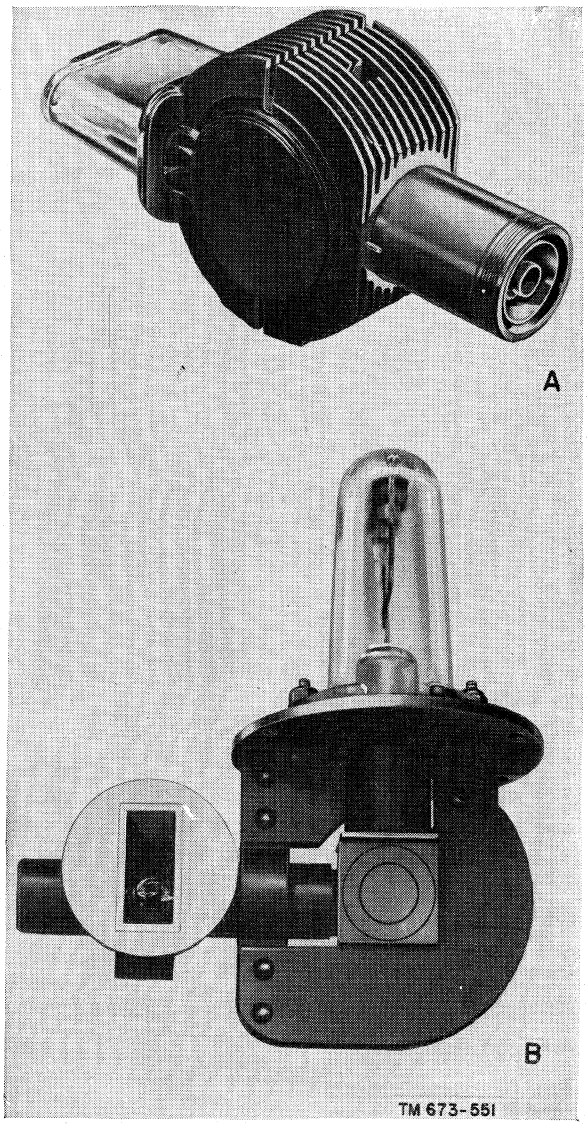


Figure 193. Typical magnetrons.

c. Magnet. The magnet of the cavity magne-

tron usually is separate from the cathode and the anode, and the term magnetron normally is applied to the cathode-anode structure and the inclosing envelope. In figure 193, the magnetron shown in A operates in the S-band and has an output connection suitable for feeding rigid coaxial cable. The magnetron in B operates in the X-band and has an output connection suitable for feeding waveguide. Figure 194 is a simplified view of the magnetron structure showing the direction of the magnetron field, parallel to the cathode. Figure 195 shows typical magnets (with keepers) used in conjunction with magnetrons. Some low-power magnetrons are constructed with the magnet as an integral part of the structure. These are called *packaged* magnetrons (fig. 196).

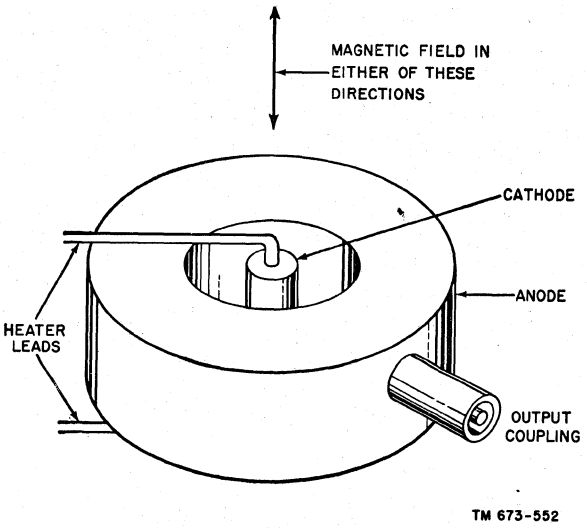


Figure 194. Direction of magnetic field.

d. Care of Magnet. A magnetron magnet usually is supplied with a *keeper*—a cylinder of soft iron which should be placed between the pole pieces of the magnet whenever the magnetron is removed. *It is necessary to resist the temptation to play with the magnet* by bringing screw drivers and pieces of metal near, since this weakens it and will affect the performance of any magnetron that subsequently is used with it. Do not wear a wrist watch when working with a magnetron magnet, for even if it is of the nonmagnetic type it probably will be damaged. The poles of the magnet normally are marked *North* and *South* and, although the magnet may be placed around the magnetron in either position, one position may be preferable to the other. If such is the case, the magnetron will be marked, and it should be inspected for a notation of this sort before putting it in service.

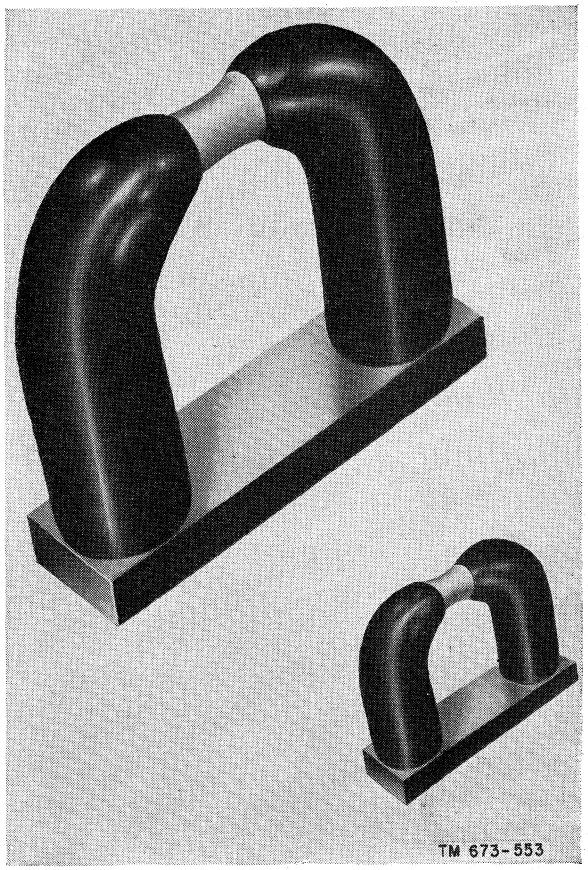


Figure 195. Magnetron magnets.

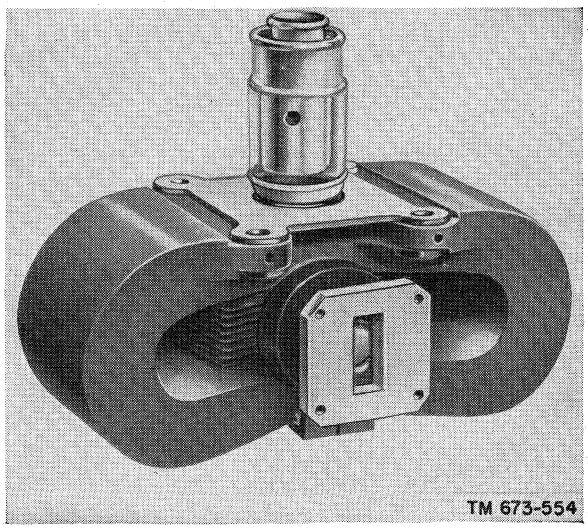


Figure 196. Packaged magnetron.

108. Theory of Operation

The magnetron depends, for its operation, on the fact that the path of an electron is deflected when the electron moves at right angles to a magnetic field. As a matter of interest, the same principle is used in every cathode-ray tube that has magnetic deflection (fig. 197).

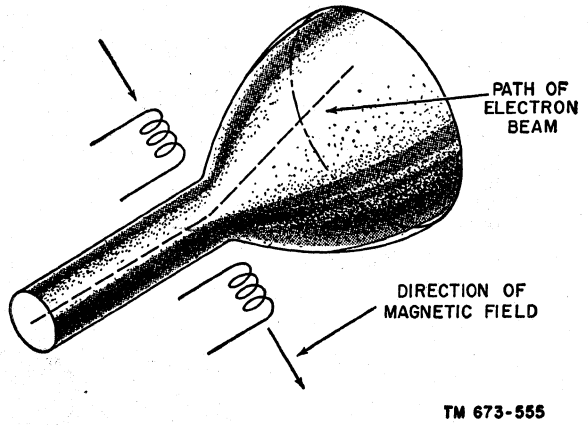


Figure 197. Magnetic deflection in cathode-ray tube.

a. Paths of Electrons in Magnetron. Aside from the cavities in the wall of the anode structure, the magnetron may be considered as a cylindrical diode with a uniform magnetic field parallel to its axis. In the absence of the magnetic field, the magnetron would behave like any other cylindrical diode; the electrons would travel radially from cathode to anode, as in figure 198. However, a warning is inserted here: DO NOT OPERATE A MAGNETRON WITHOUT THE MAGNET IN PLACE. This would ruin the magnetron. In figure 199, A shows the effect of a moderate magnetic field on the electron paths. This field, however, is much too low for normal operation. Note that the paths all are curved in the same direction by the uniform magnetic field,

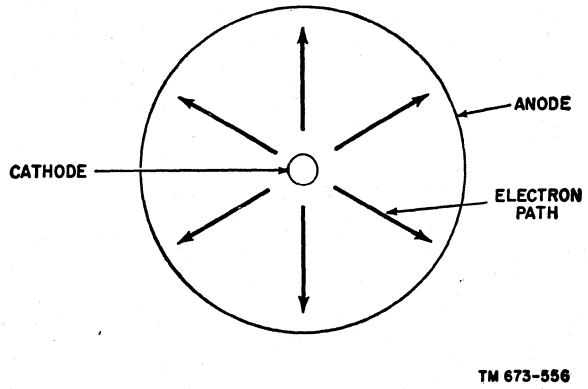


Figure 198. Electron paths without magnetic field.

which is perpendicular to the plane of the page. If this field is replaced by one in the opposite direction, the electron paths will resemble those shown in B. A stronger magnetic field, or a lower voltage between cathode and anode will lead to the paths shown in C, which are curved more sharply. When the correct relation exists between the magnetic field strength and the cathode-anode voltage, the electron paths are as those shown in A of figure 200. This is a rather idealized picture, since it neglects all interaction between the electrons and also neglects the effect of the cavities. In practice, collisions between the electrons disrupt the simple path that runs from cathode to anode and back again. The actual path spirals outward so that the electrons ultimately are captured by the anode, as shown in B.

b. Operating Conditions. Although paths of the proper sort can be obtained with any anode voltage if the appropriate magnetic field is used, or with any magnetic field if the correct anode voltage is used, the possibility of oscillation (transfer of energy from the electrons to the cavities) depends on the relations between electron velocity, which is a function of the anode voltage, spacing

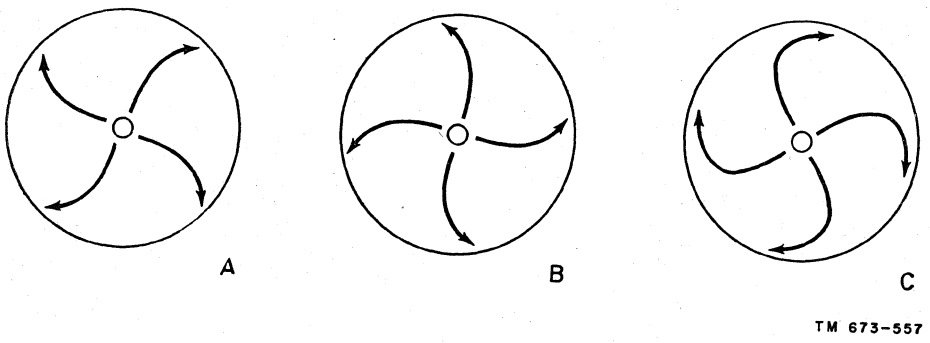


Figure 199. Effect of moderate magnetic field on electron paths.

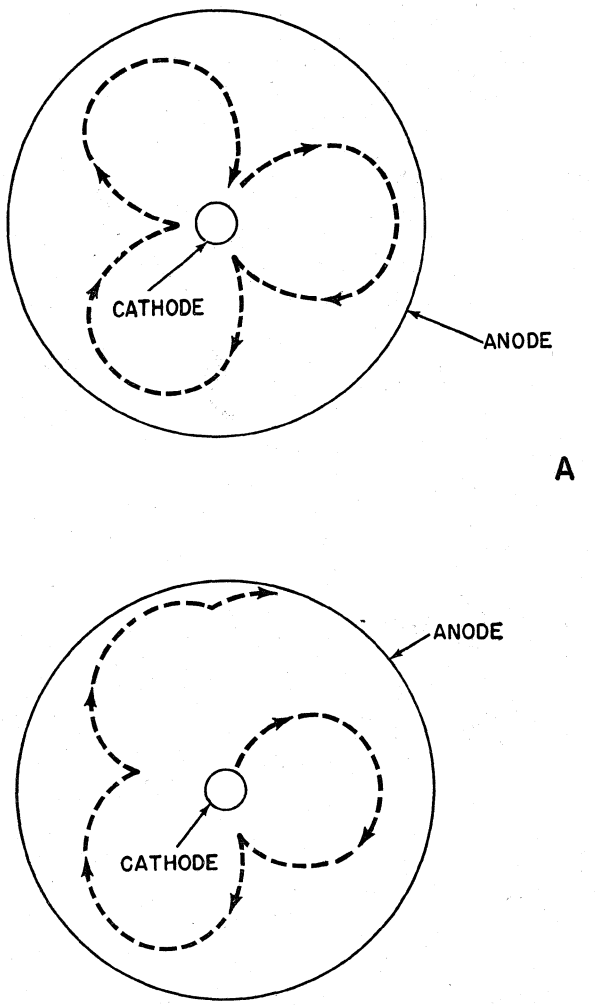


Figure 200. Electron paths with proper magnetic field.

between the cavities, and resonant frequency of the cavities. The magnetron will, therefore, oscillate over only a relatively narrow range of magnetic field strengths and anode voltages. Figure 201 shows the relation between anode voltage, anode current, magnetic field strength, and power output for a typical magnetron (2J50) operating near 9,000 mc, in the lower part of the X-band. A graph of this sort is known as a performance chart.

c. Effect of End Plates. Operation of the magnetron is not involved. Because of the magnetic field, electrons pass by the resonant cavities and transfer energy to them. As electrons slow up, they are removed from the system by the anode. The picture is complicated, although only slightly, by the presence of the end plates. Figure 202 is a simplified sketch of a magnetron, showing the end plates, which form a portion of the wall of the evacuated space. In practice, the body of the magnetron is operated at ground potential and the cathode is anywhere from 10,000 to 50,000 volts negative. In pulsed operation, negative d-c pulses are supplied to the cathode by a high-power pulse generator called a modulator. A voltage difference, therefore, exists between the cathode and the end plates, and some of the electrons wander off to the end plates without delivering much energy to the cavities. Changing the position of the end plates alters the electron paths and also, to some extent, changes the resonant frequencies of the cavities. Advantage is taken of this fact in *tunable* magnetrons, whose operating

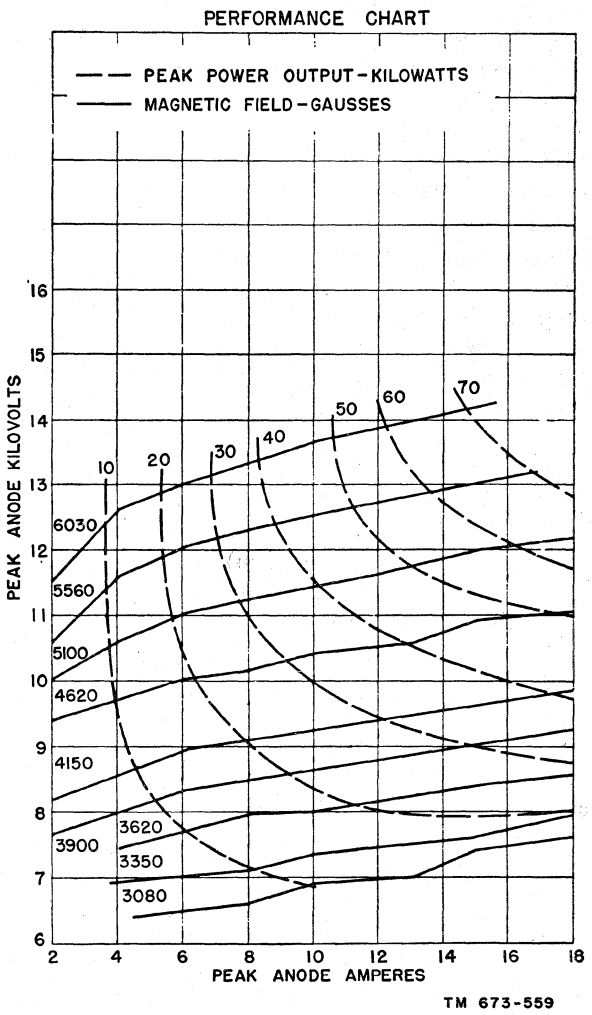


Figure 201. Variation of power output and anode current with change in magnetic field strength and anode voltage.

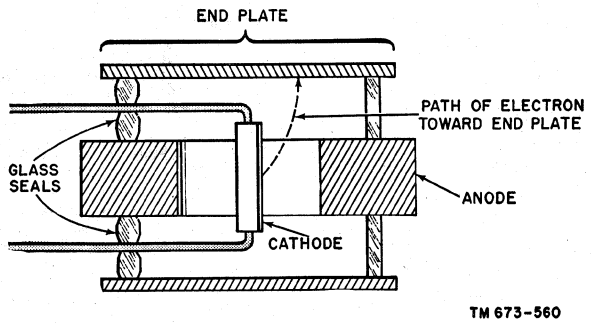


Figure 202. Effect of end plates on electron paths.

frequency can be varied by about 10 percent through distortion of the end plates.

d. Models of Operation. It is possible for a magnetron to operate in a number of different modes. In the case of a six-cavity magnetron, for example, the phase difference between adjacent cavities may be 60°, in which case the sum of the six phase-differences is 1 cycle. If the phase difference is 120°, the sum is 2 cycles; or if 180°, the sum is 3 cycles. Similarly, in an eight-cavity magnetron, the phase difference between adjacent cavities may be 45°, 90°, 135°, or 180°. The desired mode is the one with phase differences of 180°, and other modes frequently are suppressed by means of *strapping*. The strapping for a six-cavity hole-and-slot anode is illustrated in figure 203, with A showing an end view of the straps, and B showing a perspective view. The straps simply connect corresponding points on adjacent cavities. Since the straps have an electrical length of 1 half-wave, a phase difference of 180° is assured. Only one lip of each cavity is shown strapped in A, but the other lips are strapped at the other end of the anode (not shown). Other strapping arrangements are possible, of course, and all the straps may be at one end of the anode, or there may be a complete set of straps at each end. A magnetron of this sort is called a *strapped* magnetron.

e. Frequency Pulling. In view of the complex nature of the magnetron oscillator, it is not surprising to find that both frequency and power output vary when the load impedance is changed. Unfortunately, the variation in frequency for a given change in load is greatest when the load is adjusted for maximum power output from the magnetron (the input being held constant). It is, therefore, customary to adjust the load for a minimum VSWR at the magnetron rather than for maximum magnetron output. A graph of

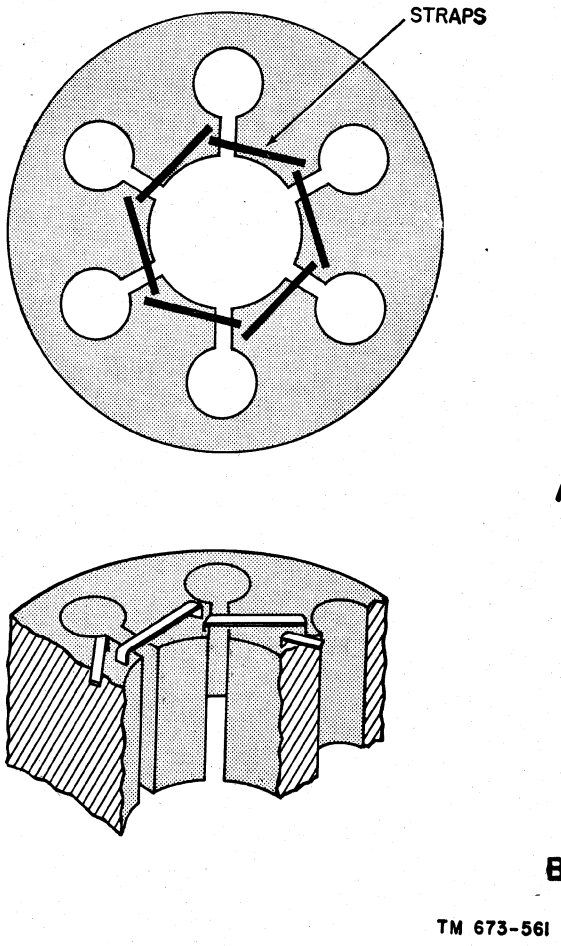


Figure 203. Strapped anode.

power output and frequency as a function of VSWR usually is called a Rieke diagram, shown in figure 204 for a 2J50 magnetron. Since the power output and frequency for a given VSWR may take on any of the values on the corresponding VSWR circle, the only way of achieving reproducible operating conditions is to tune the transmission system for minimum VSWR at the magnetron.

f. General. Magnetrons are available in sizes up to 1,000 kilowatts output for pulse service with a duty cycle of about 1/1,000 (in operation only 1/1,000 of the time). Higher duty cycles, and even continuous operation, are available in lower-powered magnetrons. In general, as the operating frequency is increased, the structure of a magnetron must be made smaller. The corresponding reduction in heat dissipation ability lowers the available power output. Most magnetrons operate with efficiencies of approximately

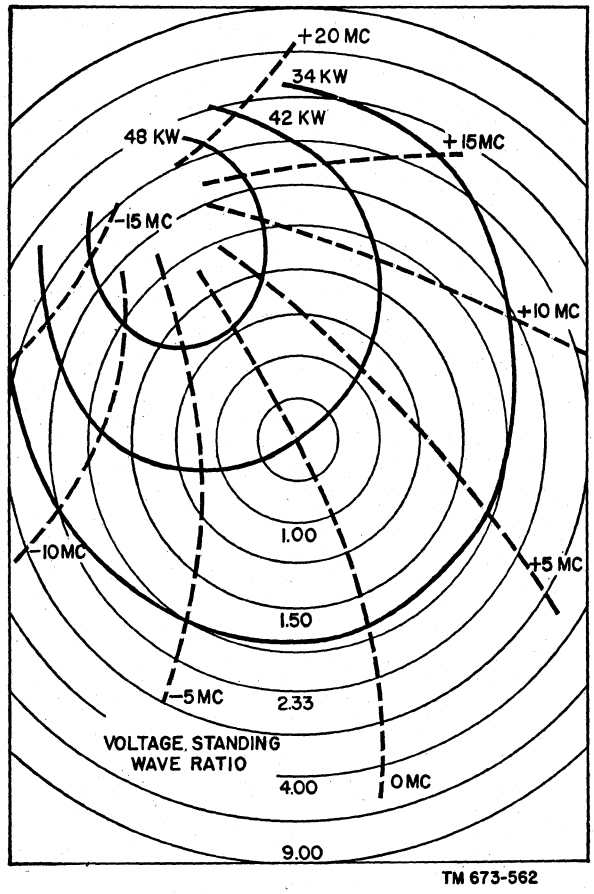


Figure 204. Rieke diagram.

35 percent, and, consequently, any reduction in dissipation capability has a marked effect on the power input that can be used, and on the power output. It is, therefore, important to make sure that blowers or other cooling systems are operating properly. Magnetrons can be modulated by variation of the anode voltage, but this results in both amplitude and frequency modulation and is generally unsatisfactory. For this reason, various forms of pulse width and pulse position modulation are used in high-power microwave systems which use magnetrons. Considering, now, the problem of stability rather than of modulation, most magnetrons are stable to better than one-half of 1 percent in frequency with variation in load. There are various methods of achieving greater stability through the use of external high-Q cavities, but these seldom are used. Microwave receivers are inherently broad-band in any event, and slight frequency variation in a microwave transmitter can be tolerated. If necessary, the receiver can be provided with an automatic frequency control system (ch. 6), so that it follows the transmitter through most variations in frequency.

109. Summary

a. Although there are a number of types of oscillators, most of these are variants of the basic feedback oscillator, which consists of an amplifier with a feedback loop from its output to its input. The operating frequency of a feedback oscillator is one for which the phase shift around the whole loop is 1 cycle. The operating level is one for which the gain around the whole loop is exactly one.

b. In most oscillators using conventional tubes, interelectrode capacitance and transit time limit the frequencies that can be reached. These limits have been raised by special construction, good examples of which are acorn tubes, lighthouse tubes, and micropups.

c. The Barkhausen oscillator is the only circuit using conventional tubes in which energy is transferred to the resonant circuit by separate electrons rather than by the electron stream as a whole (acting in a conventional amplifier).

d. Special techniques have been developed for microwave oscillators, and two classes of microwave oscillators have appeared—low-power oscillators (klystrons) and high-power oscillators (magnetrons). No high-power amplifiers are available at microwave frequencies. Both klystrons and magnetrons make use of velocity modulation, in which electron velocity in a stream of electrons is altered.

e. There are two basic types of klystron, the simple klystron, with buncher and catcher, and the reflex klystron, using only one cavity. Of these, the simple klystron is nearly obsolete. The reflex klystron may be operated in either cw (continuous wave) or pulse service and can be easily amplitude- or frequency-modulated.

f. A magnetron is, essentially, a multicavity velocity-modulated system in which electrons are forced into circular paths by means of a magnetic field. Although magnetrons may be operated cw, most of them are used in pulse service with duty cycles of approximately 1/1,000. Some magnetrons may be tuned, and all magnetrons are subject to slight variation in frequency by change in magnetic field strength or by change in anode voltage. In addition, they are subject to frequency pulling by variation in load impedance.

Amplitude and frequency modulation of magnetrons usually are not used, pulse width and pulse position modulation being preferred.

110. Review Questions

a. Is there a satisfactory method of amplifying microwave signals at low power level?

b. Is a klystron used for low- or high-power transmission?

c. What are the usual components of a microwave transmitter using a magnetron oscillator?

d. What is the chief difference between a conventional vacuum tube and a klystron or magnetron?

e. What is the basic characteristic of a feedback oscillator?

f. Explain the difference between a velocity-modulated vacuum tube and a conventional one.

g. What is the name of the graph commonly used to show the behavior of a velocity-modulated beam of electrons?

h. What is bunching?

i. What are the four basic parts of a simple klystron?

j. Of what is the beam source of a klystron composed?

k. How is the electron beam velocity adjusted in a klystron?

l. What is the purpose of the *catcher* in a klystron?

m. What is used to feed energy back from the *catcher* cavity to the *buncher* cavity in a klystron?

n. When the distance between buncher and cavity is 3VT/4K, in what mode is a klystron operating?

o. Why is it generally not feasible to operate a simple klystron in the fourth mode?

p. Why is a klystron valuable as an oscillator but of little use as an amplifier?

q. What distinguishes a *reflex* klystron from a *simple* klystron?

r. By what four means may a reflex klystron be modulated?

s. What are the three basic parts of a cavity magnetron?

t. What function has the magnet in a cavity magnetron?

u. What are the three types of cavity magnetrons?

v. May a magnetron be operated before it has come to operating temperature?

w. What is a *keeper*?

x. Why should pieces of metal be kept away from the magnet of a magnetron?

y. May a magnetron be operated without its magnet?

z. On what three factors does oscillation in a magnetron depend?

aa. What effect in a magnetron does change in the position of the end plates have?

ab. What phase differences does the desired mode in a magnetron have?

ac. What is the purpose of *strapping* in a magnetron?

ad. What is the electrical length of a strap?

ae. Is a magnetron load adjustment usually made for minimum or for maximum VSWR?

af. What is the graph of magnetron power output and frequency as a function of VSWR?

ag. Where is a magnetron transmission system tuned?

ah. What primarily limits the operating efficiency of a magnetron?

ai. Why is it important that blowers or other cooling systems attached to a magnetron system be operating properly?

aj. What is the approximate operating efficiency of most magnetrons?

CHAPTER 6

MICROWAVE RECEIVERS

Section I. RADAR

111. Introduction

Almost all radar receivers are of the superheterodyne type, shown in figure 205. Incoming signals are passed through an r-f section, which may or may not contain an r-f amplifier, to a frequency converter section. There, the signals are converted, by heterodyning with the output of a local oscillator, to a lower frequency called the intermediate frequency. The i-f (intermediate-frequency) output of the frequency converter is amplified in an i-f section and then passed to a suitable detector, which is followed by appropriate video amplifiers. Since this manual is concerned only with microwave techniques, rather than with radar techniques in general, many aspects of radar receivers are touched on only lightly in the following discussion. Further information may be found in the following manuals: TM 11–665, which deals with a-m (amplitude-modulation) equipment and techniques; TM 11–668, which deals with f-m (frequency-modulation) equipment and techniques; TM 11–672, which deals specifically with pulse techniques. The output of the radar receiver usually is shown on some sort of display equipment, discussed in TM 11–671, and also may be used to control automatic tracking equipment of the type described in TM 11–674.

112. Frequency Conversion

The heart of a superheterodyne receiver is the frequency converter, which receives signals at the operating radio frequency and delivers an output at the intermediate frequency. The frequency converter consists of a nonlinear element, or circuit, called a *mixer*, and a source of unmodulated signals, at a frequency close to that of the incoming signals, called a *local oscillator*. At vhf and lower frequencies, both the mixer and the local oscillator are conventional vacuum tubes. At medium and low frequencies, the mixer and the local oscillator sometimes are combined in a circuit which uses only a single vacuum tube. Such a circuit is known as a *converter*.

a. Mixer. The mixer may be any nonlinear circuit which is capable of operation at the frequency in use. (A nonlinear circuit is one in which the output is not proportional to the input.) When a single signal, of amplitude A and of frequency f, is passed through a nonlinear circuit, the output contains not only the input frequency, f, but also harmonics of the input, having frequencies $2f$, $3f$, $4f$, and so on, and some dc. When two signals, of amplitudes A_1 and A_2 and frequencies f_1 and f_2, are fed to a nonlinear circuit, the output contains dc and harmonics of *both* of the

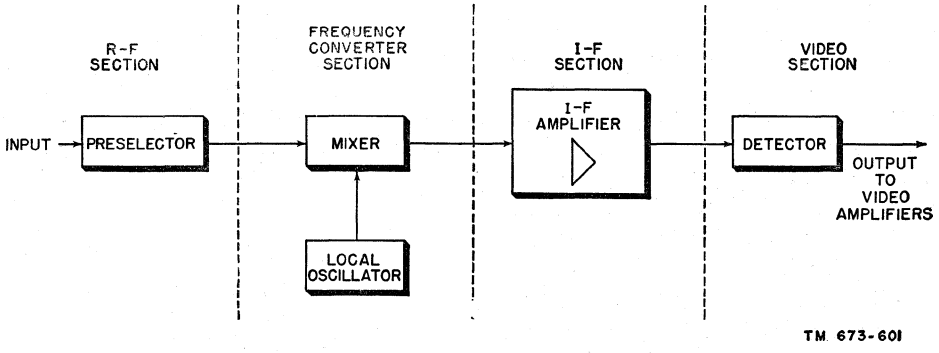


Figure 205. Superheterodyne receiver.

input signals, at frequencies $2f_1$, $3f_1$, $4f_1$, and so on *and* $2f_2$, $3f_2$, $4f_2$. In addition, the output contains signals at frequencies (f_1+f_2), (f_1-f_2), $(2f_1+f_2)$, $(2f_1-f_2)$, (f_1+2f_2), (f_1-2f_2), and so on. The fact that some of these differences lead to negative values is not important, since a signal of frequency (f_1-f_2) and a signal of frequency (f_2-f_1) are essentially the same. The important component of the output, from the standpoint of frequency conversion, is the signal having a frequency (f_1-f_2), and this signal has an amplitude proportional to (A_1A_2) if the mixer is designed properly. In an S-band receiver, for example, one of the signals may be at a frequency of 3,000 mc per second (the frequency of the incoming signal) with an amplitude P, which is proportional to the amplitude of the incoming signal. The other input to the mixer is the output of the local oscillator, which has a frequency of 2,970 mc per second and a constant amplitude, K. The desired component of the output has an amplitude proportional to (KP) and a frequency of (3,000—2,970) mc per second. This is, of course, a signal at a frequency of 30 mc per second (the intermediate frequency). Since K is constant, the signal has an amplitude proportional to that of the incoming signal. In other words, the incoming signal is reproduced (although usually at a higher level) at a frequency 2,970 mc per second lower than its original frequency. Some amplitude-modulation detectors operate in the same way, using the constant-amplitude carrier and the variable amplitude sidebands. For this reason, a mixer or converter sometimes is called a first detector, whereas the actual detector is called the *second* detector.

b. Crystal Mixers. The usual mixer used at microwave frequencies is a *crystal mixer.* This consists of a thin wire, called a *cat's whisker*, which makes contact with a small piece of silicon or germanium. The arrangement is shown in A of figure 206, with the packaged mixer shown in B. The nature of the metal, either silicon or germanium, is such that relatively few electrons are available for current carrying. Such a metal is called a semiconductor. Although it is easy to inject electrons through the cat's whisker, it is difficult to take them out in any sizable quantity because of the small area of contact. The result is that such a crystal passes current easily from the metal to the cat's whisker and with difficulty in the opposite direction. Current flow is in the opposite direction from electron flow; current in a vacuum tube flows from anode to cathode, whereas

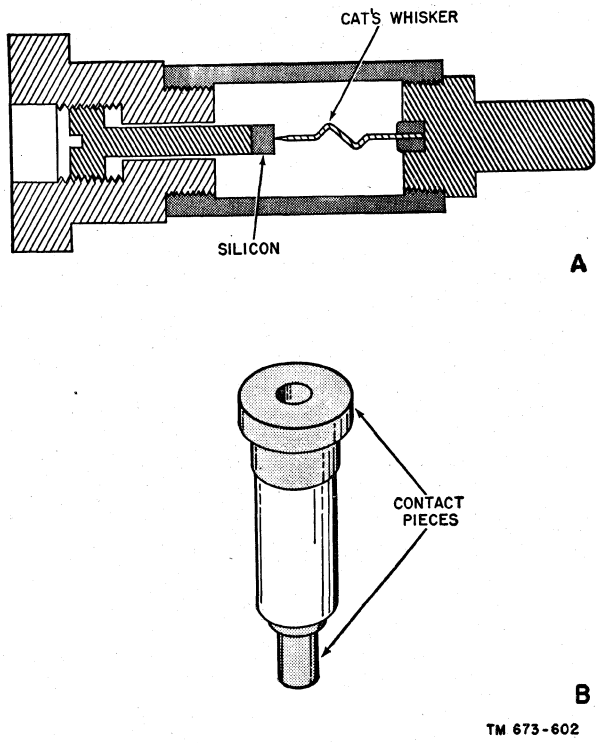


Figure 206. Crystal mixer.

electrons travel from cathode to anode. The current-voltage relationship for a crystal mixer is shown in figure 207. For applied voltage in the backward direction, the current is proportional to the voltage, and it is *very small.* For a large applied voltage in the forward direction, the current is approximately proportional to the voltage, but for a small forward voltage, the current is proportional to the square of the applied voltage. The curved portion of the characteristic, where

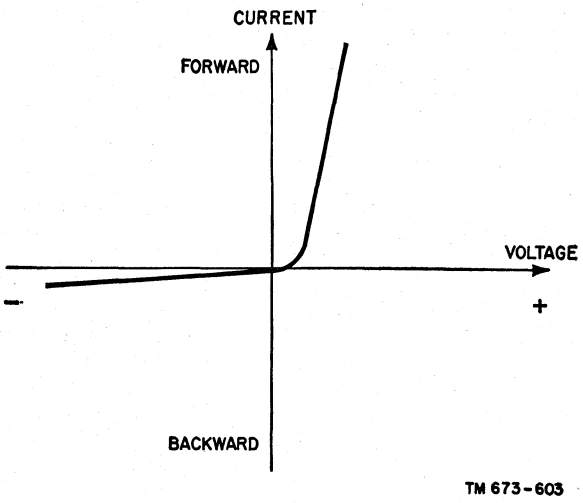


Figure 207. Voltage-current relationship in crystal mixer.

the voltage-current relation is nonlinear, is used for mixing. The general construction of a waveguide frequency converter (local oscillator and mixer) is shown in A of figure 208, with the equivalent circuit in B. The method of mounting the crystal mixer is shown clearly in the figure. A coaxial line frequency converter (used only on S-band) is shown in A of figure 209, and the equivalent circuit is illustrated in B.

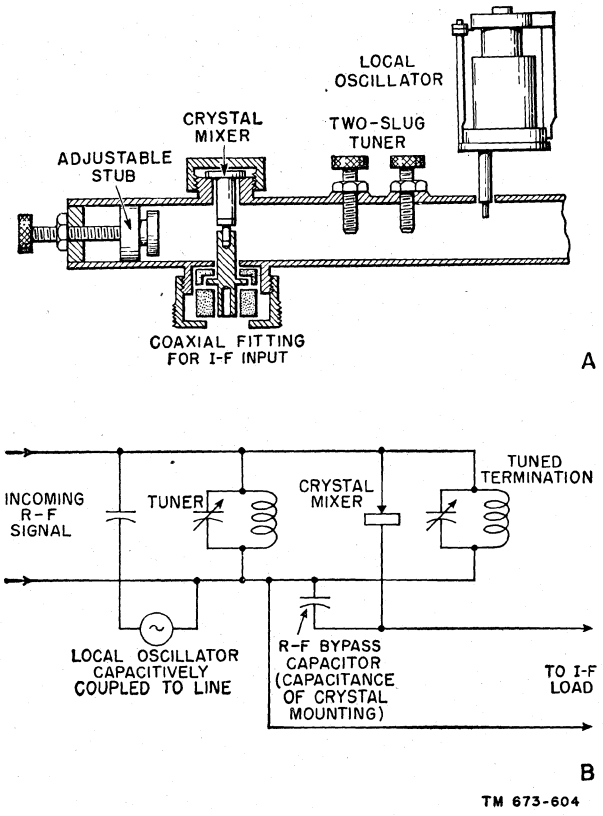


Figure 208. Waveguide frequency converter.

c. Care of Crystal Mixers. Since a crystal mixer is damaged easily by overloading, it usually is supplied in an envelope of thin metal—frequently lead foil. The envelope serves the double purpose of protecting the mixer from moisture and shielding it from strong electric fields which might burn it out. When installing a crystal mixer, *touch the crystal mount with one finger just before inserting the crystal.* This discharges any static electricity that may have been generated, possibly by friction between shoe soles and floor and guarantees that there will not be a short pulse of high voltage across the crystal.

d. Other Mixers. Reference to figure 207 shows that the voltage-current relation for a crystal mixer is the same as that of a diode (the crystal mixer often is called a crystal rectifier or crystal diode). The same sort of crystal is used as a microwave detector, or rectifier, in connection with slotted line sections and directional couplers which are described in chapter 3. Any other diode may be used as a mixer, although pentagrid mixers are used at lower frequency because they amplify the signal in addition to shifting its frequency. The difficulty with most diodes is that the anode-cathode capacitance completely bypasses the diode at microwave frequencies. Aside from crystal mixers, the only mixer of any importance for operation at microwave frequencies is the lighthouse diode type 559. This is shown in figure 210. The two signals, r-f input and local oscillator input, are fed through the diode to the i-f load. The section of transmission line consisting of the anode sleeve and the cathode sleeve is a tuning impedance which resonates the diode capacitance, at the r-f frequency. The section of line consisting of the cathode sleeve and the outer conductor acts as an r-f bypass capacitor in parallel with the i-f load. Mixers of this sort, however, seldom are used. The crystal mixer is both smaller and simpler.

e. Local Oscillators. The usual local oscillator is a reflex klystron. Less frequently, and only on the S-band, a simple klystron is used. Occasionally, also on the S-band, a lighthouse local oscillator is used together with a lighthouse diode mixer. The lighthouse oscillator is discussed in chapter 5.

f. Frequency Converter Assemblies. A complete frequency converter assembly, consisting of a TR switch, local oscillator, crystal mixer, and appropriate tuners, frequently is made as a single unit. A shutter seals off the crystal. When the equipment is turned on, the shutter is withdrawn by a solenoid after the TR switch is energized. The crystal mixer, therefore, is protected at all times, either by the shutter or by the TR switch, against damage by excessively strong signals.

g. Radar Beacons. In many applications, particularly in airborne operation, a single radar is used both for search and as a navigational instrument. Navigation information is obtained by interrogating *radar beacons* or *racons.* A racon is nothing more than a transmitter and a receiver connected in such a manner that the transmitter sends out a distinctive signal, or a series of signals, whenever the receiver picks up an interrogating signal (fig. 211). The racon receiver *listens* on

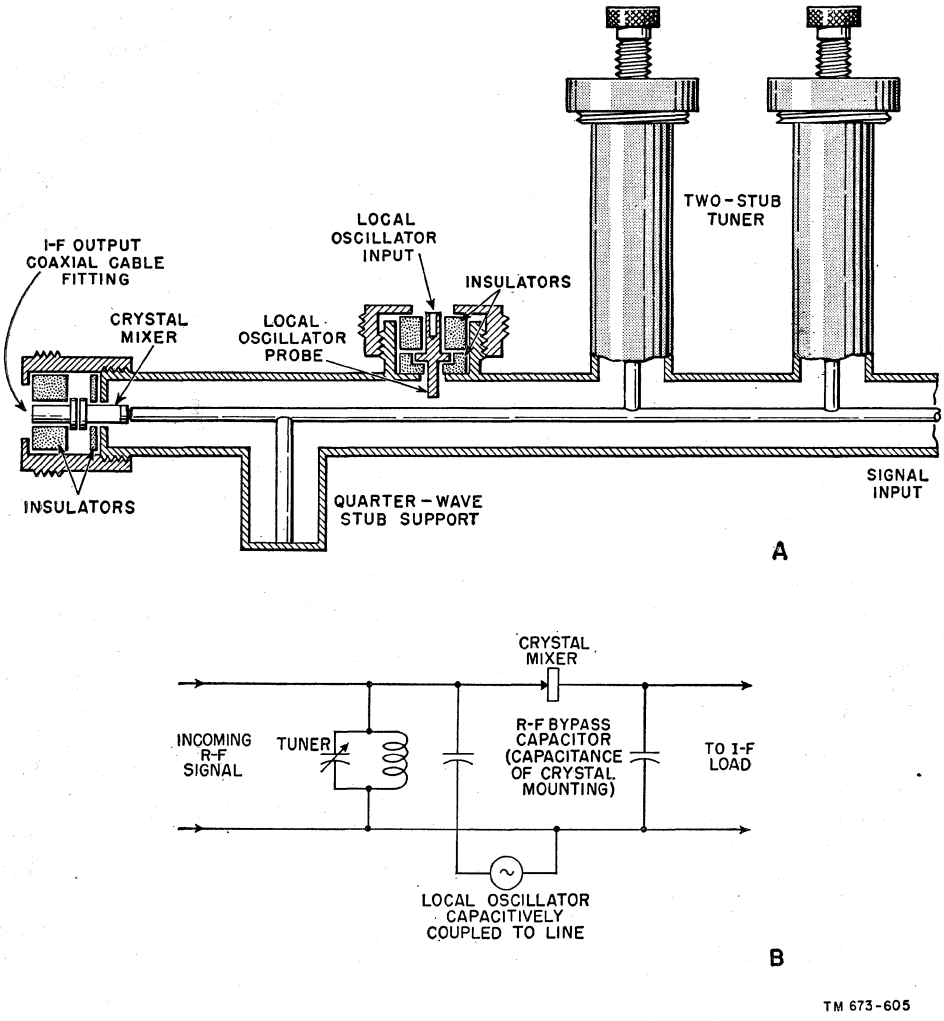


Figure 209. Coaxial line frequency converter

the frequency normally used by the search radar. The racon transmitter, however, *replies* on another frequency so that its signal easily may be distinguished from the normal radar echoes. Many search radar receivers are required to operate on either of two frequencies, one for search operations, the other for racon navigation.

h. Search-and-Beacon Frequency Converter Assembly. Frequency converter assemblies sometimes are designed for operation on either of two frequencies. They differ from the simpler converter assembly described above by having two local oscillators. One is tuned for normal search operation, the other is tuned for reception on the racon band. Tuning from one frequency to the other is, then, a simple matter of turning one klystron off and turning the other on. For simplicity in tuning, a two-frequency assembly of this sort usually is provided with two crystal mixers and two TR switches as well as two sets of tuning devices.

113. Preselection

In a superheterodyne receiver, the r-f amplifier, if used, is called a preselector. At frequencies below the microwave region, it is customary for high-performance receivers to employ one or two stages of preselection in order to improve image rejection and receiver sensitivity, and to eliminate local oscillator radiation. At microwave frequencies, however, preselectors are used only in the S-band, and then only rarely.

a. Image Rejection. Consider a superheterodyne receiver designed to receive signals at a frequency f_1, which uses an intermediate fre-

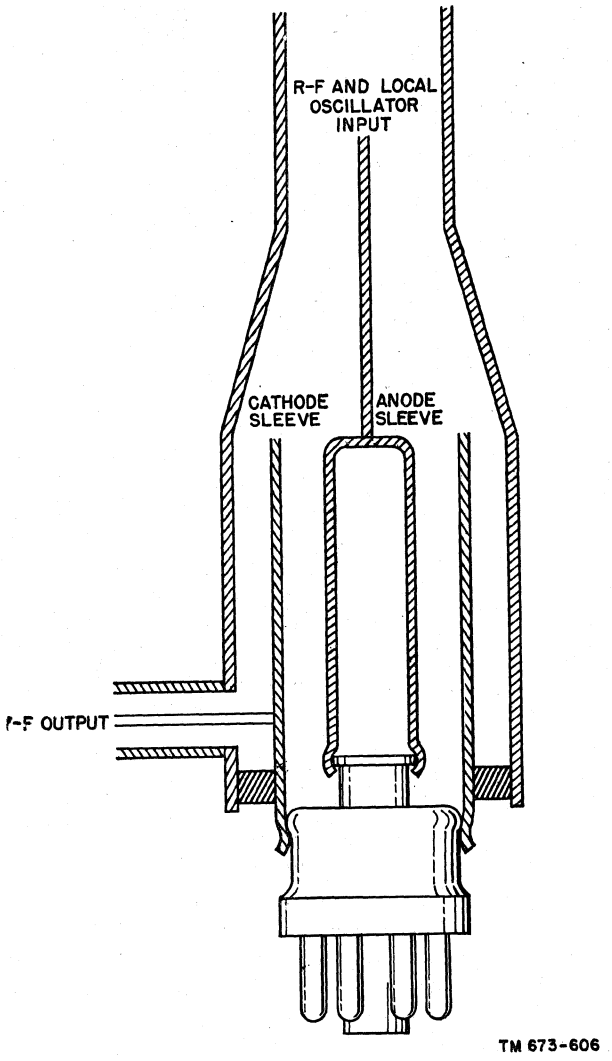


Figure 210. Lighthouse diode.

quency f_2. If the local oscillator frequency is to be lower than that of the incoming signals, the oscillator will be tuned to a frequency (f_1-f_2), as in A of figure 212. The difference between the frequency of the incoming signals and that of the local oscillator is, then, $f_1-(f_1-f_2)$, or f_2. The mixer will, however, deliver an output at the intermediate frequency f_2 if there is an incoming signal at the *image frequency* (f_1-2f_2), for the difference, $(f_1-f_2)-(f_1-2f_2)$, is also equal to the intermediate frequency. The local oscillator is also operated at a frequency higher than that of the incoming signal, in which case the local oscillator is tuned to a frequency (f_1+f_2), as in B. Here, the image frequency is (f_1+2f_2). There is, then, an image frequency which is separated from the desired operating frequency by *twice* the intermediate frequency. Since the mixer will deliver an output at the intermediate frequency with an input at the desired frequency *or* with an input at the image frequency, all portions of the receiver after the mixer will treat signals at the image frequency in precisely the same way that they treat signals at the desired operating frequency. The only way of rejecting signals at the image frequency is by means of tuned circuits *ahead* of the mixer. The mixer input normally is tuned to the operating frequency f_1, and this offers some image rejection. The insertion of additional tuned circuits associated with one or two stages of amplification, at the r-f frequency f_1, offers greatly increased rejection of signals at the image frequency.

b. Receiver Sensitivity. In a superheterodyne receiver, the output level for a given signal input

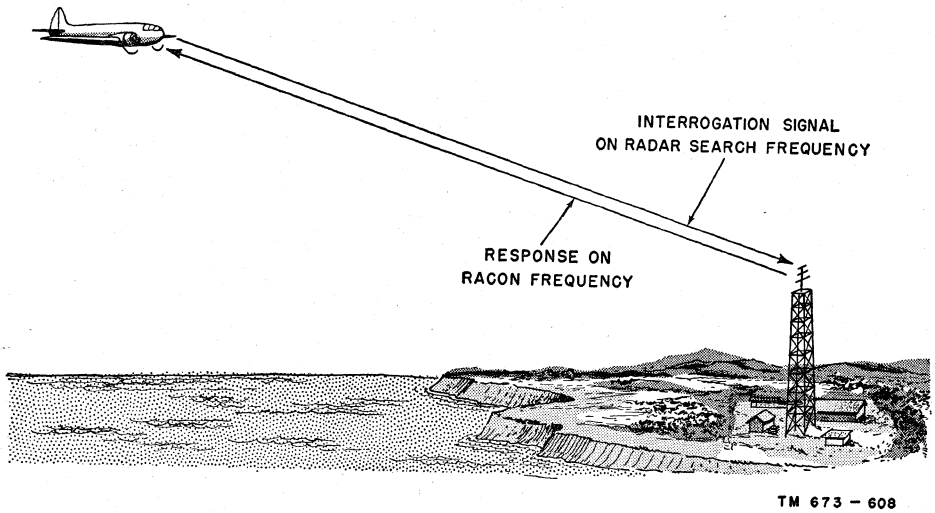


Figure 211. Beacon operation.

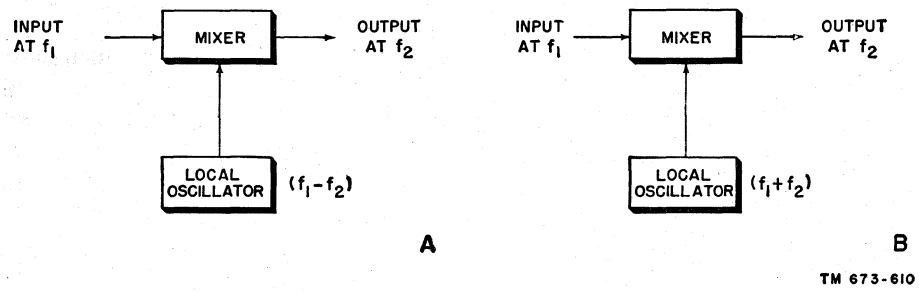


Figure 212. Frequency converter.

is set by the over-all receiver gain—by the amplification in the preselector, mixer, intermediate-frequency amplifier, and the audio or video amplifier. The selectivity, or ability to reject signals at frequencies just outside the desired bandpass, is determined by the number of tuned circuits in the intermediate-frequency amplifier. Image rejection is determined by the number of tuned circuits in the preselector. The receiver sensitivity, however, depends on the *equivalent noise input* at the first stage of the receiver. Every electrical circuit is troubled by noise. In a circuit like a crystal mixer, the noise results from random motion of the electrons in the conductor. In a vacuum tube, the noise results from roughness in the electron stream as well as from random electron motion in the conductors of the grid circuit. On this account, every receiver has some output even when no signal is received from the antenna. Suppose that, for a given receiver, this output averages 10 volts when the receiver is connected to a 100-ohm load. This is, of course, equal to 1 watt. If, now, an input signal of 5 microvolts raises the output voltage to 14 volts, which indicates a power output of about 2 watts, the signal power output and the noise power output are equal. The receiver noise, therefore, may be considered as the equivalent of a 5-microvolt signal, since it develops the same power at the receiver output. Although there are ways of detecting signals that are well below the level of the background noise, it is clear that the minimum usable signal is related to the equivalent noise input, since any reduction in the noise level will allow detection of weaker signals. A preselector improves the sensitivity of a receiver (in general) because it has a lower equivalent noise level than a mixer, although this is not always true. Assume, for example, that a given receiver has an equivalent noise input of 5 microvolts, and that a preselector with a gain of 10 and an equivalent noise input of 2 microvolts is inserted ahead of the receiver. With the preselector, the noise at the mixer will be the original noise, equivalent to a 5-microvolt signal, plus the preselector noise which, after amplification, is equivalent to a 20-microvolt signal. Then, at the mixer, the level of the amplified preselector noise is 4 times as great as that of the mixer noise, and the *power* in the amplified preselector noise is 16 times as great as the power in the mixer noise. The mixer noise is, therefore, relatively unimportant and the equivalent noise input of the receiver is that of the preselector, or 2 microvolts.

c. Local-Oscillator Radiation. Whenever a mixer is connected directly to the transmission line from an antenna, it is possible for some of the signal from the local oscillator to travel along the same transmission line *to* the antenna, where it is radiated and may cause interference with other equipment. The insertion of an r-f amplifier in the line between the antenna and the mixer eliminates this possibility, since the amplifier will carry signals *from* the antenna but not *to* the antenna.

d. Microwave Preselectors. Preselectors are not used at microwave frequencies in the X- and K-bands, since there is no practical method of amplifying signals at these frequencies. It is possible to use a simple klystron (not a reflex klystron) as an r-f amplifier at S-band frequencies, but this seldom is done because the equivalent noise input of the klystron is higher than that of the crystal mixer. The only effective preselector is a lighthouse tube amplifier and even this is seldom used, because it offers little advantage.

e. Grounded-Grid Amplifiers. When a lighthouse tube is used as an r-f amplifier at S-band frequencies, it is used in a *grounded-grid* circuit. The grid is grounded, the input signal is developed

between cathode and ground, and the output signal appears between plate and ground. Consider the amplifier shown in A, figure 213. This is a conventional circuit, except for the fact that the grid-cathode capacitance, C_{gc}, and the grid-anode capacitance, C_{ga}, are shown on the drawing. Suppose, for a moment, that the grid voltage is raised slightly. It is, of course, necessary to charge capacitor C_{gc} by delivering current from the driving source. If the anode potential remains fixed, capacitor C_{ga} must be charged from the same source, since the plate load is a high impedance for signal frequencies. The effective input capacitance of the tube is, therefore, the sum $(C_{gc}+C_{ga})$. That is, the circuit behaves exactly as though a capacitor of value $(C_{gc}+C_{ga})$ were connected from grid to ground. *In fact,* when the grid is made more positive, the anode moves in the negative direction A times as far, where A is the amplification of the stage. The voltage change across the grid-anode capacitance is, therefore, $(1+A)$ times as great as the voltage change at the grid, and it requires $(1+A)$ times as much charging current as it would if the plate voltage did not change. It behaves, in short, like a capacitance A times as big as C_{ga} connected between grid and ground. The equivalent input capacitance of the circuit is, then $(C_{gc}+C_{ga}+AC_{ga})$ Suppose, now, that the grid is grounded, as shown in B, figure 213. When the cathode is made more positive, it is still necessary to charge capacitor C_{gc}. The grid-anode capacitance, however, now is charged entirely through the plate load circuit. The cathode anode capacitance must be charged, but this is much smaller than the grid-anode capacitance. In addition, if the cathode is made more positive, the plate also moves in the positive direction. With a gain of 1, both plate and cathode move through the same voltage change, and no charging current is required for C_{ca}. With greater gain, a negative charging current is required, so that the current to charge C_{gc} is obtained, in part, from C_{ca}. The resulting equivalent input capacitance is, then, only $(C_{gc}+C_{ca}-AC_{ca})$, and may be reduced nearly to zero. A typical grounded-grid lighthouse amplifier is shown in A of figure 214, with the equivalent circuit in B. The amount of noise in a receiver depends largely on the noise originated by nonuniformity in the electron flow in the first tube. If r-f amplification is used, this tube is the first r-f tube. Otherwise, with a crystal mixer, it is the first i-f tube. Instead of giving the equivalent input noise, it is sometimes convenient to compare the equivalent noise power input of the receiver with the noise power contributed by the antenna, which results from random motion of the electrons in space near the antenna. A *noise figure* of 1 would indicate that the receiver contributed no noise of its own. In practice, noise figures of 35 to 100 are usual. Except at the very lowest microwave frequencies, the noise figure for a receiver with an r-f amplifier is likely to be higher than that for the same receiver with a crystal mixer. For this reason, preselectors seldom are used even at the low microwave frrequencies where the required r-f amplifiers can be constructed. As stated above, preselectors are not used on the X- and K-bands because no r-f amplifiers are available.

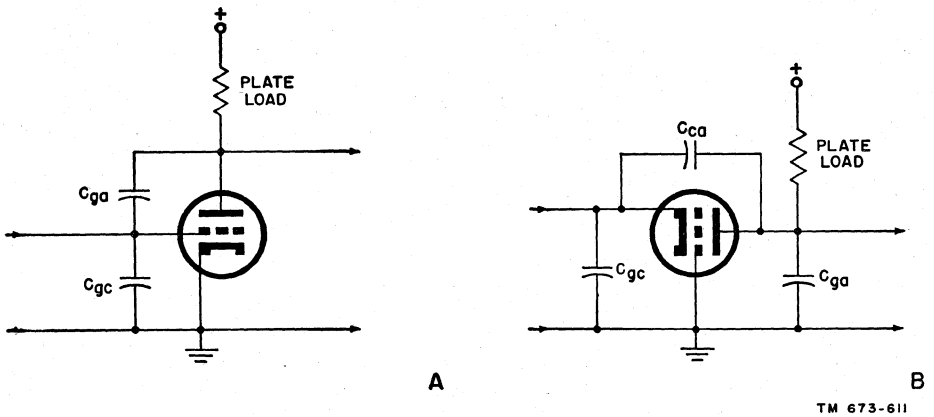


Figure 213. Conventional and grounded-grid amplifier circuits.

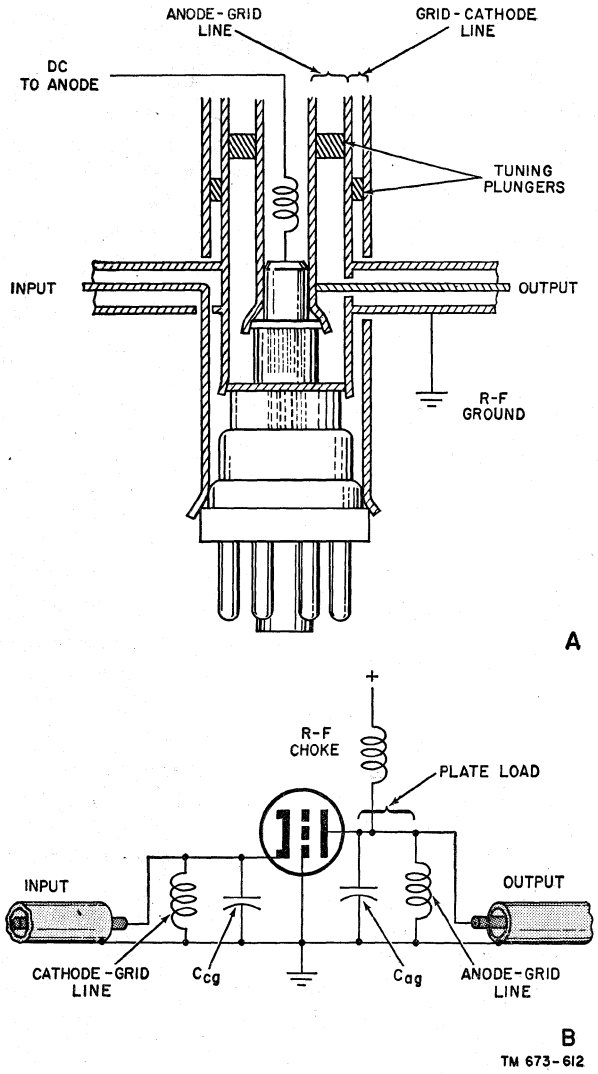


Figure 214. Grounded-grid lighthouse amplifier.

114. Intermediate-Frequency Amplification

The intermediate-frequency amplifier of a radar receiver is much like any other i-f amplifier. The only specific requirement is that of bandwidth. For conventional, or pulse-type radar, it is customary to have a bandwidth of $3/d$ mc per second in the intermediate-frequency amplifier, where d is the pulse width in microseconds. Most search radars use pulse widths of approximately 1 microsecond, and adequate bandwidth can be obtained with an intermediate frequency of 30 mc per second. Tracking radars, in general, use shorter pulses and usually operate at an intermediate frequency of 60 mc per second. These two frequencies have become more or less standard for vhf and microwave operation, and other intermediate frequencies seldom are encountered. Radar beacon receivers, which must pick up all interrogating pulses whether or not the interrogating transmitters are operating at the proper frequency, normally use an intermediate frequency of 60 mc per second in order to achieve greater bandwidth.

a. Bandwidth and Noise. As the bandwidth of a receiver is increased, the noise output also increases. The equivalent noise input, therefore, is proportional to the bandwidth. In a search radar, the basic problem is the detection of targets through the background of noise. The bandwidth is, consequently, reduced as far as possible in order to minimize the equivalent noise input. There is, however, a limit below which the bandwidth cannot be reduced without impairing the performance of the receiver. With a bandwidth of $1/d$, signal pulses and noise pulses are similar in appearance, and it is difficult to decide whether a small receiver output is the result of noise or of a weak signal. The customary bandwidth of $3/d$ passes more noise power than does a bandwidth of $1/d$, but it also makes echoes more distinguishable from noise. With tracking radars, the basic problem is the precise determination of range and, in some cases, echo amplitude. Since tracking radars operate at relatively short ranges and, therefore, with relatively strong signals, the bandwidth frequently is increased to $5/d$ in order that echo signals may have sharper fronts. This, of course, permits the target range to be determined with greater precision.

b. Coupling From Mixer to I-F Amplifier. The usual method of coupling to the intermediate-frequency amplifier is by means of coaxial cable from the mixer. The coaxial line is terminated in its characteristic impedance by a link inductively coupled to the tuned circuit at the grid of the first i-f amplifier tube (fig. 215). Other coupling methods also are used. A meter, with a bypass capacitor shunting it, frequently is inserted in

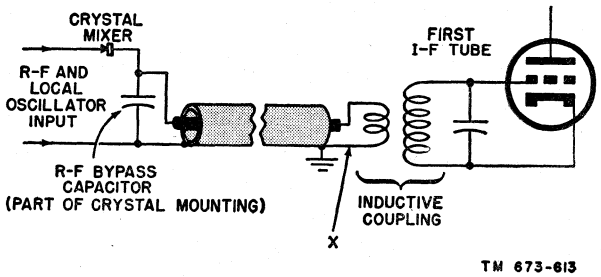


Figure 215. Coupling to i-f amplifier.

the circuit at the point marked X in order that the crystal current may be read. The meter serves as a convenient indicator for adjustment of the level of the local oscillator. Usual crystal currents are something less than 1 milliampere.

115. Detection and Amplification

The method of handling the output of the intermediate-frequency amplifier differs with different types of radar receivers. Detection and subsequent treatment of the signal in a number of types of receivers are described briefly in *a* through *f* below.

a. Pulse-Type Radar.

(1) In a pulse-type radar system, the transmitter sends out a series of short pulses of energy. These are reflected from any targets in view, and echo pulses are returned to the radar receiver. The receiver, accordingly, *listens* for echo pulses during the intervals between the transmitted signals. The range of a given target is determined by noting the time interval between transmission of an outgoing pulse and arrival of the corresponding echo from the target in question. The direction of a given target is determined by noting the direction from which the echo pulses arrive. The simplest method of determining target direction involves the use of a highly directional antenna, for both transmission and reception. The echo from a given target, then, is seen only when the antenna is oriented so that the target lies in the beam. A more complicated method of determining target direction in a single plane involves rapid shifting of the radiation pattern back and forth through a small angle. Figure 216 shows the operation of such a system, known as a *lobe-switching* system. The radar alternately transmits and receives with

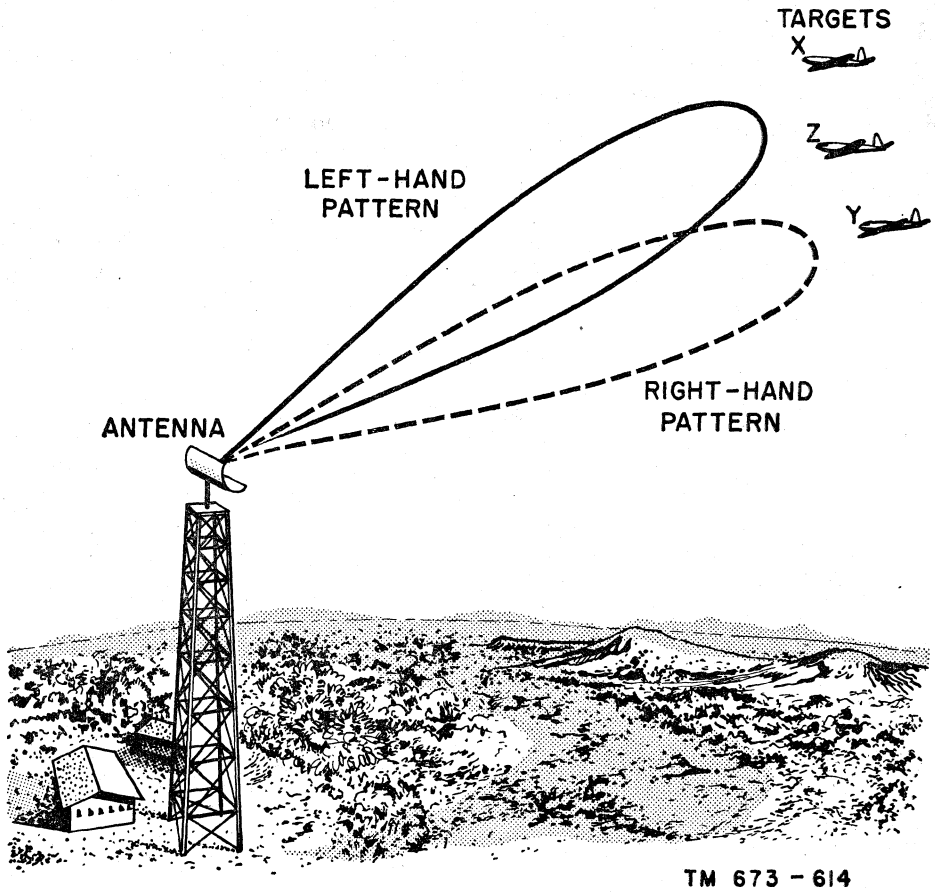


Figure 216. Lobe-switching system.

the left-hand pattern and with the right-hand pattern, the pattern being switched several times each second. Obviously, a target at X will return a stronger echo when the radar is operating on the left-hand pattern than when it operates on the right-hand pattern. For a target at Y, the situation is reversed, and the stronger echo is returned during operation on the right-hand pattern. For a target at Z, in the center of the pattern system, an echo of constant strength is returned, since the target is just as far from the center of the left-hand pattern as from the center of the right-hand pattern.

(2) For operation in three dimensions, a *conical-scanning* system is used. Here the beam is swept around the surface of a narrow cone, as shown in figure 217. A target on the axis of the cone will return an echo of constant strength. A target that is off the axis of the cone will return an echo which is amplitude-modulated at the conical scanning frequency, usually 30 cycles. The amount of modulation shows how far the target lies from the axis of the scanning cone, and the phase of the modulation, relative to the phase of the scanning motion, shows the direction in which the target lies. In both lobe-switching and conical-scanning systems, it is necessary to consider the percentage of modulation rather than the actual modulating signal. A nearby target slightly off the center of the pattern will return a strong echo with a moderate amount of modulation. A distant target far off the center of the pattern will return a weak echo with, possibly, the same *amount* of modulation but with a higher *percentage* of modulation. The modulation itself may be used as an indication of how far the target lies from the center of the pattern only after the average echo strength has been brought to some standard level.

b. Search Radar (Pulse-Type). In a search radar, it is desirable to examine all echoes. The output of the intermediate-frequency amplifier is, therefore, detected in a conventional amplitude-modulation detector. Following this, the signal is amplified in a video amplifier and sent to a suitable display system where it may be observed by an operator. The display system normally involves a cathode-ray tube and is one of the several types of displays described in TM 11–671.

c. Tracking Radar (Pulse-Type). In a tracking radar, the output of the intermediate-frequency amplifier is detected, as in the case of a search radar, and amplified in a video amplifier to some convenient level. At this point, the desired echo is selected by means of a *gated amplifier* controlled by a *range gate.* An amplifier of this sort is shown in figure 218. The tube normally is cut off by the negative bias applied to grid 1. At an appropriate time after transmission of the outgoing pulse, the range gate signal arrives at grid 1 and momentarily overcomes the negative bias. The tube does not conduct, however,

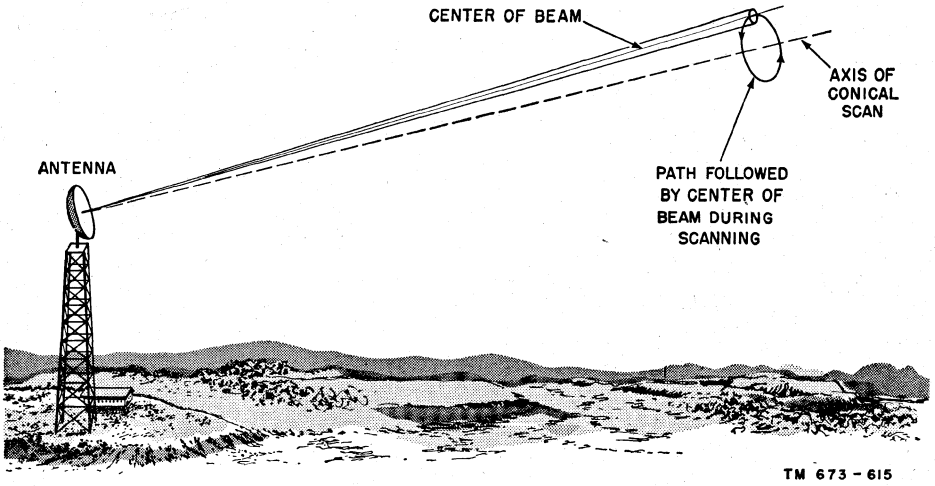


Figure 217. Conical scan system.

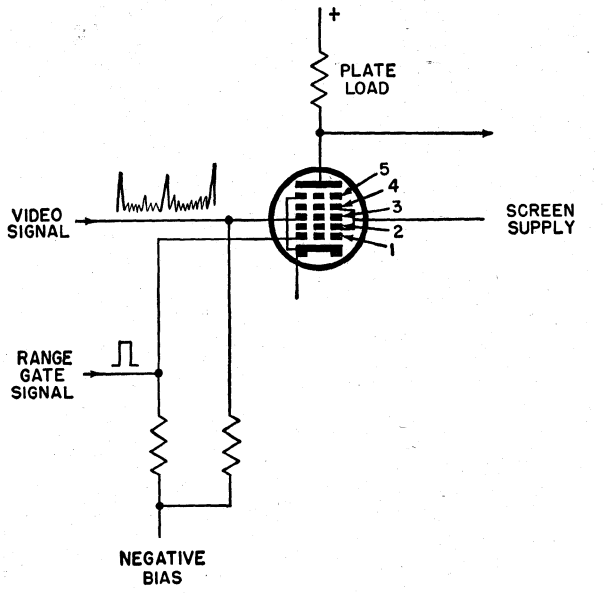


Figure 218. Gated amplifier.

unless a signal arrives at grid 3 during the time that the gate is *opened* by the range gate signal. Echoes from the target under observation thus are passed by the range gate, and all others are rejected. At the same time, the range of the target under observation is determined easily by noting the time interval between transmission of the outgoing signal and generation of the range gate signal. The gated signal, after amplification, is detected again in a low-frequency detector, which recovers the modulation on the echo pulses. The modulation then may be used to control the direction in which the antenna points. The control is accomplished by means of servo systems, which are discussed in detail in TM 11–674. Either the gated video signal, which contains echoes from the target under observation, or the ungated video signal, which contains all of the echoes received, may be sent to a display system for observation by an operator.

d. Automatic-Tracking Range Gate. In some radars, the range gate is held on the desired echo by an operator. In others, an automatic-tracking range gate is used. In the automatic-tracking system, three gated amplifiers are required. One is controlled by an *early gate*, one by a *signal gate*, and one by a *late gate*. The time relationships of the gates is shown in figure 219. The early gate opens first; shortly after this the signal gate opens, and finally the late gate opens. The gates then close in the same order. As shown in figure 219, there is a slight amount of overlap between the early gate and the late gate. In operation, the signals from the early gate amplifier and those from the late gate amplifier are compared. If the group of range gates is set at a range slightly greater than the target range, almost all of the echo will be passed by the early gate and very little will be passed by the late gate. The output of the early gate amplifier then will be greater than that of the late gate amplifier, and the resulting difference is used to reduce the range at which the gates are set. If the range is too small, the output of the late gate amplifier will be greater than that of the early gate amplifier, and an appropriate correction can be made. The range setting thus is changed automatically until the output of the early gate amplifier is equal to that of the late gate amplifier. Under this condition, the output of the signal gate amplifier contains echoes from the desired target. As the range changes, the three gates are shifted automatically so that the target continues to appear in the signal gate.

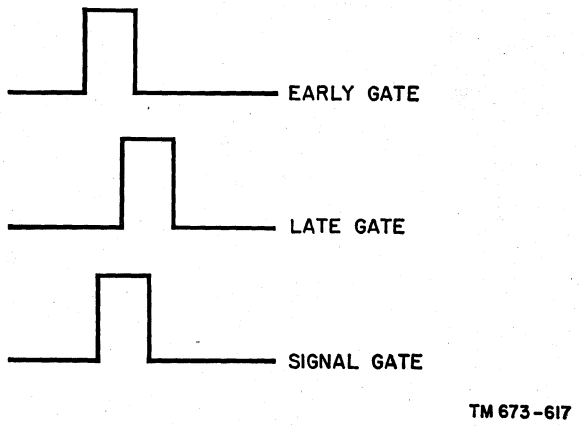


Figure 219. Gates for automatic-tracking range gate system.

e. Frequency-Modulated Radar.

(1) A specialized type of radar, of which the low-altitude absolute altimeter is a good example, makes use of frequency-modulated transmission rather than pulsed transmission. Consider a transmitter operating in the neighborhood of 500 mc per second and frequency-modulated in the manner shown in figure 220. At every instant, the transmitter frequency is changing at the rate of 1,000 mc per second, in one direction or the other. If, now, the ground, or any other prom-

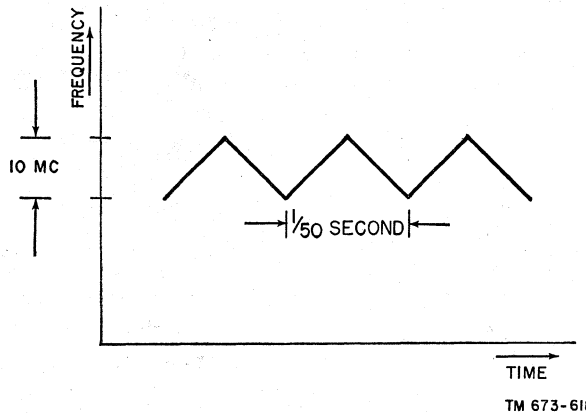


Figure 220. Frequency modulation of f-m radar transmitter.

inent target, is approximately 500 feet from the transmitter, the echo will arrive after a time interval of 1 microsecond. During most of the time (all of the time except just before and just after the direction of the frequency change is reversed), the frequency of the received signal will differ from that of the transmitted signal by 1,000 cycles per second. The range to the target, or the distance to the ground, then can be determined by measuring the frequency difference. This can be done directly by using the transmitter as the local oscillator in the receiver. The intermediate frequency then will be the desired frequency; the difference in frequency between the transmitted and the received signals will be equal to the intermediate frequency. In practice, however, this is difficult, because of the low level of the received signal.

(2) One alternative is the scheme shown in figure 221. The received signal is picked up on an antenna which is shielded from the transmitting antenna. The signal then is sent to one of a pair of mixers, the other mixer being fed by a low-level signal from the transmitter. Both mixers make use of the same local oscillator, and both input frequencies are shifted by the same amount. The two mixer outputs, which are amplified in identical intermediate-frequency amplifiers, therefore have the same frequency difference that exists between the transmitted and the received signals. After amplification to a suitable level, the two signals from the intermediate-frequency amplifiers are sent to a third mixer. The difference frequency is selected at the output of this mixer by a low-pass filter, and is sent to a frequency meter which

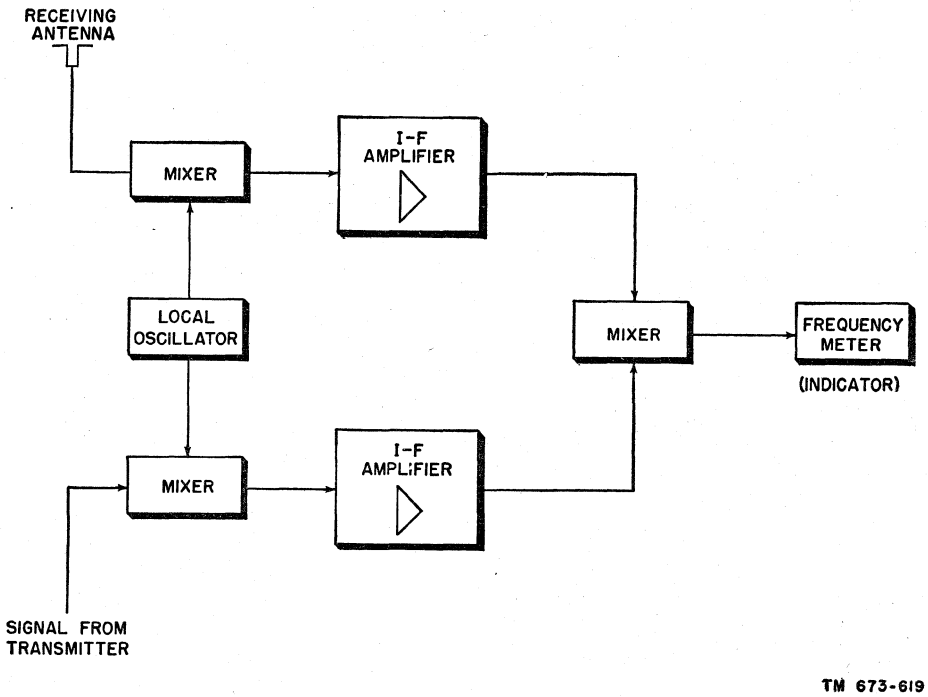


Figure 221. F-m radar receiver.

reads directly in range. Frequency-modulated radar of this sort is useful when only one target at a time is in the combined transmission-reception pattern. Although the example given is a vhf system, the same technique can be used at microwave frequencies.

f. Continuous-Wave Radar. Another type of radar, c-w radar, makes use of receiving equipment similar to that used with f-m radar. The transmitter, however, is not modulated at all but transmits a continuous signal at constant amplitude and frequency. Suppose, for example, that there is a target at a range of 1,000 wavelengths of the transmitted signal. The received signal, having traveled a distance of 2,000 wavelengths, will lag the transmitted signal with a phase difference of 2,000 cycles. If, now, the target is moved toward the transmitter to a distance of 999 wavelengths, the phase difference will be reduced to 1,998 cycles. Suppose, now, that the target moves toward the transmitter at a rate of 1 wavelength per second. The phase difference then decreases at the rate of 2 cycles per second. If the transmitted signal goes through f cycles in any second (where f is the transmission frequency), the received signal must go through $(f+2)$ cycles in the same second in order to reduce the phase difference by 2 cycles during the 1-second time interval. The velocity of the target toward the transmitter or away from it is, then, given directly by the frequency meter, a frequency difference of 2 cycles per second, corresponding to a velocity of 1 wavelength per second. One difficulty with the system as described is the impossibility of telling whether the target is approaching or going away. This may be solved by heterodyning the transmitter signal, as it comes from its i-f amplifier, with, say, a 1,000-cycle signal. The resulting difference frequency is, then, 1,000 cycles per second lower than the intermediate frequency which corresponds to the transmitted signal. If this difference-frequency signal is sent to the final mixer in place of the output of the transmitter signal i-f amplifier, the frequency meter will show a frequency of 1,000 cycles per second when the target is stationary. A velocity of 1 wavelength per second toward the transmitter then will be shown by a frequency reading of 1,002 cycles per second, and a velocity of 1 wavelength per second away from the transmitter will be shown by a reading of 998 cycles per second. Note that the system shows *range rate* only. If the target moves along a circle the center of which is at the transmitter, the range will remain constant and the system will report no target motion.

116. Automatic Gain and Frequency Control

An agc (automatic gain control) system in a conventional communications receiver is designed to maintain a fixed average signal level at the output of the intermediate-frequency amplifier, even though the signal level at the receiver input varies over a wide range. Simple control circuits of this sort are seldom used in radar receivers, however, although most radar receivers do have some form of agc in conjunction with afc (automatic-frequency control), which keeps the receiver tuned to a selected incoming signal. Some representative circuits of this type are described in *a* through *d* below.

a. Conventional Automatic Gain Control System. The ordinary agc system used in communications receivers is very simple. It involves nothing more than a detector, fed by the output of the i-f amplifier, and a low-pass filter (A of fig. 222). A portion of the output of the amplifier output is fed to a conventional diode. This rectifies the i-f signal and develops a negative voltage across the load resistor that is proportional to the amplitude of the signal. The i-f bypass capacitor passes the i-f component in the rectifier output, and the voltage across the load resistor is nothing more than the envelope of the i-f signal. This voltage then is passed through a low-pass filter which removes any variation caused by amplitude modulation at audio or higher frequencies. The output of the filter, with the diode connection as shown in A, is, then, a negative voltage that is proportional to the *average* signal level at the output of the i-f amplifier. This negative voltage is used to reduce the gain of the i-f amplifier, usually by shifting the grid bias of each tube in the negative direction. Suppose, now, that an input signal level of 10 microvolts develops a signal of 1 volt at the output of the i-f amplifier. What increase in the input level is required to raise the i-f output level to 1.1 volts? With 1 volt of agc voltage, the receiver gain, from the receiver input to the i-f amplifier output, is 100,000. This, of course, is the ratio of 1 volt to 10 microvolts. With an agc voltage of 1.1 volts, the receiver gain may be reduced by a factor of 10. The new gain then will be only 10,000

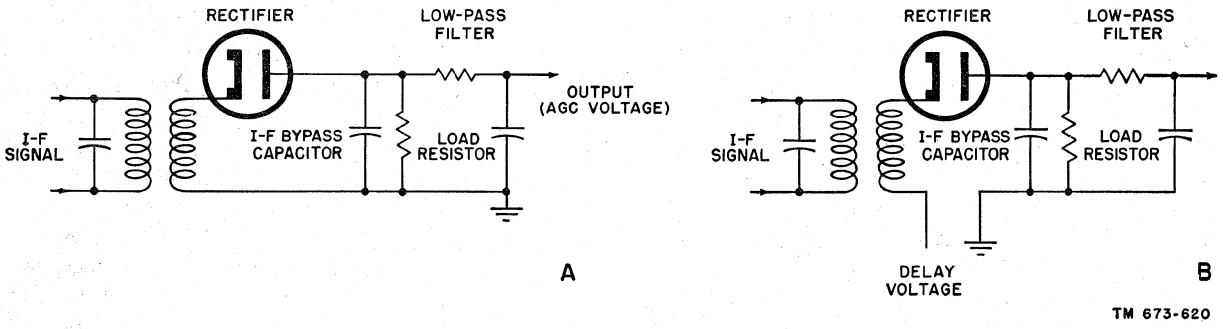


Figure 222. Simple agc systems.

and the required input signal will be 1/10,000 of 1.1 volts, or 110 microvolts. With the system as described, there is only a 10-percent change in the i-f output voltage, but the input signal level is made 11 times as great (a change of 1,000 percent). In many cases, it is desirable to retain the maximum receiver gain for small input signals. The simple agc system then is replaced by a *delayed* agc system, such as that shown in B. Here, a positive *delay voltage* is applied to the diode in series with the i-f output. This prevents the diode from conducting when the i-f signal level is low. There is, therefore, no agc voltage developed until the i-f signal level exceeds the bias voltage. Simple agc systems of this type are not suitable for use in radar receivers, except in rare cases where it is desirable to have the *strongest* echo control the gain of the receiver. Ordinarily, this is undesirable, since the receiver gain then is reduced to bring the strongest echo to some standard level. Other echoes, in this case, are at lower levels and may be overlooked.

b. Gated Automatic Gain Control System. It is customary, in tracking radars, to use a form of automatic gain control called *gated* agc. The echo passed by the range gate is sent to a rectifier, and the resulting voltage is passed through a low-pass filter as in a conventional agc system to remove the error modulation, which results from error in the position of the radar antenna. This system differs from a simple agc system only in the fact that the agc voltage is developed from a *single* series of echoes—those from the target under observation. As explained above, the gated echo must be brought to some standard level in order to have any fixed relation between the amplitude of the error modulation and the extent of the antenna position error. This is carried out by the gated agc system.

c. Sensitivity Time Control System. Although it is not feasible to use an agc system with a search radar receiver, because the strongest echo received would control the gain of the receiver, it is usual to employ an stc (*sensitivity time control*) system, which varies the receiver gain in such a way that echoes from distant targets are amplified more than those from nearby targets. Consider, for the moment, two similar targets, one at a range of 10 miles, the other at twice that range. Since the power level of the transmitted signal falls off as the square of the range, the power delivered to the nearer target is four times that delivered to the farther target. The nearer target, being excited at a power level four times as great as that which excites the farther target, returns an echo that is four times as powerful as the echo returned by the farther target. The echo from the nearer target, traveling only half as far as that from the farther target, suffers only one-fourth as much attenuation in traveling to the receiver. The result is that the echo from the nearer target is received at a power level 16 times as great as that of the echo from the farther target. The power level is proportional to the square of the amplitude, and the echo from the nearer target is at a *voltage* level four times as great as the voltage level of the echo from the farther target. *Both echoes will be displayed at the same amplitude* if the echo from the farther target is amplified four times as much as the echo from the nearer target. This is particularly desirable, since the nearer target otherwise would be displayed as a strong echo that easily might distract the operator and cause him to overlook the echo from the farther target. Uniform signals will be displayed for targets of a given type at all useful ranges, if the receiver gain is made to increase with time in the proper way. The required variation in gain is provided by an stc system. The heart of such a system is a circuit that develops an increasing voltage which is